[modified]

PTO/SB/21 (09-04)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission

Application Number	09/618,178
Filing Date	18 July 2000
First Named Inventor	LINCOLN, Stephen E.
Art Unit	1637
Examiner Name	FREDMAN, Jeffrey Norman
Attorney Docket Number	13151-2

ENCLOSURES *(Check all that apply)*

- [ ] Fee Transmittal Form
- [X] Fee Attached $500.00
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD
- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [X] Appeal Communication to TC (Appeal Notice, Brief, Reply Brief)
- [ ] Proprietary Information
- [ ] Status Letter
- [X] Other Enclosure(s) (please Identify below):

  Receipt Acknowledgement Postcard

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Kalow & Springut LLP		
Signature	J. David Ellett, Jr.		
Printed name	J. David Ellett, Jr.		
Date	14 November 2006	Reg. No.	27,875

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below:

Signature	J. David Ellett, Jr.		
Typed or printed name	J. David Ellett, Jr.	Date	14 November 2006

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Before the Board of Patent Appeals and Interferences

In re Appeal regarding Patent Application of	:	
Applicants : LINCOLN, Stephen E. and KNAPP, Michael R.	:	Examiner: FREDMAN, Jeffrey N.
Serial No. : 09/618,178	:	Art Unit 1637
Filed : 18 July 2000	:	
Title : Automatic Genotype Determination	:	

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

14 November 2006

Commissioner for Patents
Post Office Box 1450
Alexandria, Virginia 22313-1450

APPEAL BRIEF

I. Real Party in Interest

The real party in interest with respect to the application on appeal; application No. 09/618,178; is the assignee of record: Beckman Coulter, Inc.; a corporation of Delaware having a place of business at 4300 North Harbor Road, Fullerton, California 92834-3100.

II. Related Appeals and Interferences

The subject application on appeal is:

a continuation of application No. 09/088,820, filed 2 June 1998 ("the '820 application");

which is a continuation of application No. 08/362,266, filed 22 December 1994 ("the '266 application"), now patent 5,762,876;

which is a continuation-in-part of application No. 08/173,173 ("the '173 application"), filed 23 December 1993, now abandoned;

which is a continuation-in-part of application No. 07/775,786, filed 11 October 1991 ("the '786 application"), now patent 6,004,744;

which is a continuation-in-part of application No. 07/664,837, filed 5 March 1991 ("the '837 application"), now patent 5,888,819; and

said application No. 08/173,173, filed 23 December 1993, is also:

a continuation-in-part of application No. 08/162,397, filed 6 December 1993 ("the '397 application"), now abandoned; and

a continuation-in-part of application No. 08/155,746, filed 23 November 1993 ("the '746 application"), now patent 5,518,900; and

a continuation-in-part of application No. 08/145,145 ("the '145 application"), filed 3 November 1993, now abandoned.

The subject application for which the present appeal is pending on appeal has not been the subject of an interference or any judicial proceeding.

The attorneys for the applicants submit that each of the '173 application, the '266 application, and the '820 application as filed provides full support for the currently pending

claims of the subject application. A review of the attorneys' office files of the '173 application, the '266 application, and the '820 application revealed no appeal, interference, or judicial proceeding involving any of the three applications.

It is submitted that none of the '786 application, the '837 application, the '397 application, the '746 application, and the '145 application provides full support for any currently pending claim of the subject application and that any appeal, interference, or judicial proceeding involving any of '786 application, the '837 application, the '397 application, the '746 application, and the '145 application would have no bearing on the Board's decision in the pending appeal. For the sake of completeness only, it is noted that the '837 application was involved in two interferences: interference No. 103,562 and interference No. 103,739. As pointed out above, the '837 application ultimately issued as United States patent 5,888,819.

III. Status of Claims

Claims 1 – 74 (Cancelled without prejudice)

Claims 75 and 76 (Finally rejected; on appeal)

Claim 77 (Cancelled without prejudice)

Claims 78 – 82 (Finally rejected; on appeal)

Claims 83 and 84 (Cancelled without prejudice)

Claim 85 – 87 (Finally rejected; on appeal)

Claims 88 – 90 (Cancelled without prejudice)

Claims 91 – 98 (Finally rejected; on appeal)

Claim 99 (Cancelled without prejudice)

Claim 100 (Finally rejected; on appeal)

Claim 101 (Cancelled without prejudice)

Claim 102 (Finally rejected; on appeal)

Claims 103 – 105 (Cancelled without prejudice)

Claims 106 – 115 (Finally rejected; on appeal)

IV. Status of Amendments

No amendment to the claims or specification to the subject application has been filed subsequent to the final Office Action of 12 October 2005 on appeal.

The following papers have been filed in connection with the application subsequent to the 12 October 2005 Office Action on appeal in addition to the submission by mail with certificates of mailing dated 10 April 2006 of a notice of appeal and a petition for an extension of time under 37 CFR 1.136(a) to respond to the Office Action through 10 April 2006.

A pre-appeal brief request for review with respect to the application was submitted to the Patent and Trademark Office on 10 April 2006 by mail with a certificate of mailing along with the notice of appeal and the petition for an extension of time. A notice of panel decision from the pre-appeal brief review was issued by the review panel on 11 May 2006 holding that the application remained under appeal assertedly because there was at least one actual issue for appeal.

An information disclosure statement under 37 CFR 1.97(d) together with copies of three citations and a petition to extend time under 37 CFR 1.136(a) to file an appeal brief through 14 November 2006 was submitted by mail to the Patent and Trademark Office on 13 October 2006 with certificates of mailing.

V. Summary of Claimed Subject Matter

The subject application has three independent claims: claims 75, 96, and 106; each of which has been finally rejected in the Office Action of 12 October 2005 on appeal. Claims 75, 96, and 106 are summarized in turn below, with references to the page and line numbers of the specification of the application as filed and to the drawings as filed by reference characters, in accordance with the requirements of 37 CFR 41.37(v). References to the specification of the application below are in the format [*page No. 1:line No. 1 – line No. 2*] or [*page No.1:line No. 1 – page No. 2:line No. 2*].

Independent claim 75 is directed to a method of determining the genotype of a subject at a locus within genetic material obtained from a biological sample from the subject [2:2-4, 7-9; 4:28-5:3; 5:19-21].

The method of claim 75 comprises the steps of reacting the material at the locus to produce a first reaction value indicative of the presence of a given allele at the locus [2:9-12] and forming a data set including the first reaction value [2:12-13].

The method of claim 75 includes the further step of establishing a distribution set of probability distributions, including at least one distribution, associating hypothetical reaction values with corresponding probabilities for each genotype of interest at the locus [2:13-17].

The method of claim 75 also includes the step of applying the first reaction value to each pertinent probability distribution to determine a measure of a conditional probability of each genotype of interest at the locus [2:17-20].

Finally, the method of independent claim 75 includes the step of determining the genotype based on the data obtained from the step of the preceding paragraph of applying the first reaction value to each pertinent probability distribution to determine a measure of a conditional probability of each genotype of interest at the locus [2:20-21].

Independent claim 96 is directed to a method of associating with a sample of genetic material from a subject one of a predetermined plurality of genotypic classes defined with respect to a genetic locus sited in the genetic material [2:2-4, 7-9; 13:5-14; 13:25-14:2] together with a corresponding confidence measure [9:28-10:1; 15:1-13]. Each genotypic class identifies either a possible genotype for the subject defined with respect to the genetic locus or a failed-experiment condition [9:6-8; 13:25-14:2; 17:11-17; 19:3-4] and each genotype is defined by the identity of one or more alleles defined with respect to the genetic locus[9:6-8; 13:5-8; 13:25-14:2; Figure 8, 5th column].

The method of claim 96 comprises the step of carrying out one or more allele-sensitive reactions on the genetic material of the sample at the genetic locus to obtain a plurality of quantitative allele-indicative reaction values [2:9-12, 22-25; 11:7-9; 12:10-16]. Each allele-indicative reaction value is indicative of the likely presence or absence of a particular allele defined with respect to the genetic locus [2:9-12, 22-25; 18:9-12]. The plurality of reaction values corresponding to the sample are assembled as a reaction-value data point [12:10-16, 18-24; 13:1-5; Figure 3].

The method of claim 96 also comprises the step of obtaining with respect to each of the genotypic classes corresponding reaction-value data-point conditional-probability-measure distribution information providing, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class [2:13-20; 2:27-3:2; 6:1-5; 13:9-14:5; Figures 4-7].

The method of claim 96 comprises the further step of evaluating for each of the genotypic classes the corresponding reaction-value data-point conditional-probability-measure distribution information with respect to the allele-indicative reaction values of the reaction-value data point corresponding to the sample to obtain for each of the corresponding genotypic classes a reaction-value data-point conditional probability measure of the reaction-value data point given the genotypic class [13:16-23; 14:15-27].

The method of claim 96 further comprises the step of establishing with respect to each genotypic class a genotypic-class conditional probability measure of the genotypic class given the reaction-value data point corresponding to the sample from the reaction-value data-point conditional probability measure of the reaction-value data point given the genotypic class to obtain a set of genotypic-class conditional probability measures of the respective genotypic classes given the reaction-value data point [2:17-20; 6:5-9; 13:16-23; 14:15-27].

The method of claim 96 also includes the step of selecting a highest-probability genotypic-class conditional probability measure of a genotypic class given the reaction-value data point having a highest probability value from the set of genotypic-class conditional probabilities to identify a most-likely genotypic class corresponding to said highest-probability genotypic-class conditional probability measure for association with the sample [9:28-10:1; 15:1-5].

Finally, the method of independent claim 96 includes the step of establishing a confidence measure for the association of the most-likely genotypic class with the sample. The confidence measure is established from values of the genotypic-class conditional probability measures of the respective genotypic classes given the reaction-value data point corresponding to the sample [15:1-14; Figure 8, column 6].

Independent claim 106 is directed to a method of associating with a sample of genetic material from a subject (i) one of a predetermined plurality of genotypic classes defined with respect to a genetic locus sited in the genetic material [2:2-4, 7-9; 13:5-14; 13:25-14:2] and (ii) a confidence measure for the association of said genotypic class with the sample [9:28-10:1; 15:1-13]. Each genotypic class identifies either a possible genotype for the subject defined with respect to the genetic locus or a failed-experiment condition [9:6-8; 13:25-14:2; 17:11-17; 19:3-4]. Each genotype is defined by the identity of one or more alleles defined with respect to the genetic locus [9:6-8; 13:5-8; 13:25-14:2; Figure 8, 5th column].

The method of claim 106 comprises the step of carrying out one or more allele-sensitive reactions on the genetic material of the sample at the genetic locus to obtain at least two quantitative allele-indicative reaction values [2:9-12, 22-25; 11:7-9; 12:10-16]. Each allele-indicative reaction value is indicative of the likely presence or absence of a particular allele defined with respect to the genetic locus[2:9-12, 22-25; 18:9-12]. The reaction values corresponding to the sample are processed to form a reaction-value data point [12:10-16, 18-24; 13:1-5; Figure 3].

The method of claim 106 further comprises the step of associating one of the genotypic classes with the sample using the reaction-value data point corresponding to the sample to define a sample genotypic class [2:17-21; 4:28-5:10; 6:5-11; 14:15-15:5; 17:1-6].

The method of claim 106 includes the step of obtaining with respect to each of the genotypic classes corresponding reaction-value data-point conditional-probability-measure distribution information providing, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the hypothetical reaction-value data point given the genotypic class [2:13-20; 2:27-3:2; 6:1-5; 13:9-14:5; Figures 4-7; 17:1-6].

The method of claim 106 also includes the step of evaluating the reaction-value data-point conditional-probability-measure distribution information corresponding to the genotypic class associated with the sample with respect to the reaction-value data point corresponding to the sample to obtain a reaction-value data-point conditional probability measure of the reaction-value data point given the sample genotypic class [13:16-23; 14:15-27; 15:5-14].

Finally, the method of claim 106 comprises the step of establishing a confidence measure for the association of the sample genotypic class with the sample using the reaction-value data-point conditional probability measure of the reaction-value data point given the sample genotypic class [15:5-14; Figure 8, column 6].

VI. Grounds of Rejection to be Reviewed on Appeal

VI.*a*) Whether claims 75, 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, 95, 96 through 98 inclusive, 102, 106 through 109 inclusive, and 112 through 115 inclusive were unpatentable under 35 U.S.C. §103(a) over a publication by Kimpton *et al.* in *PCR Methods and Applications*, volume 3, pages 13 through 22 (August 1993) ("the Kimpton *et al.* publication") in view of a publication by Ledwina *et al.* in *Biometrics,* volume 36, pages 160 through 165 (1980) ("the Ledwina *et al.* publication") and further as assertedly motivated in view of a publication by Jeanpierre in the *Annals of Human Genetics*, volume 56, page 325 through 330 (1992) ("the Jeanpierre publication"). (Claim 77 was identified as a pending claim in the Office Action on appeal; however, claim 77 was previously cancelled without prejudice in a Reply to a Final Office Action filed 27 January 2005.)

VI.*b*) Whether claims 75, 76, 78 through 82 inclusive, 85 through 87 inclusive, 91 through 98 inclusive, 100, 102, and 106 through 115 inclusive are unpatentable under 35 U.S.C. § 103(a) over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication, and further in view of published International Patent Application WO 92/15712 to Goelet *et al.* ("the Goelet *et al.* '712 published international application").

VI.*c*) Whether claim 94 is unpatentable under 35 U.S.C. § 112, second paragraph, with respect to whether the expression "including such amplification by a polymerase chain reaction or a ligase chain reaction" in the expression "assaying for the given allele using genetic bit analysis, allele-specific hybridization, or allele-specific amplification, including such amplification by a polymerase chain reaction or a ligase chain reaction" recited in the claim is vague and indefinite and lacks antecedent basis.

VI.d) Whether the pending claims of the application are entitled the benefit of the 23 December 1993 filing date of claimed parent application 08/173,173.

VII. Argument

VII.*a*) The Rejections Under 35 U.S.C. § 103(a)

VII.*a.1*) Background Technology and Terminology

Basic concepts of Mendelian genetic analysis underlie the terminology and the interpretation of genetic data in the science of genetic analysis as practiced in recent times, including data obtained in connection with the automatic genotype determination method disclosed and claimed in the subject patent application. A brief review of background in the principles of Mendelian genetic analysis may assist the Board of Patent Appeals and Interferences in recognizing the errors of claim interpretation and reference analysis which we submit are evidenced in the Office Action on appeal, as discussed below.

Basic principles of Mendelian genetic analysis are nicely summarized in Table 2-2 on page 26 of the book *An Introduction to Genetic Analysis, Fifth Edition*, by Griffiths *et al.* (W. H. Freeman, 1993) ("the Griffiths *et al.* textbook"). Excerpts from the Griffiths *et al.* textbook, including pages 19 through 42, have been made of record in the subject application and are included in the evidence appendix accompanying the present brief. The Griffiths *et al.* textbook was a standard introductory textbook on genetic analysis that was evidently publicly available as of the 23 December 1993 filing date of application No. 08/173,173, a claimed parent of the subject application, which parent, it is submitted, provides full support for the pending claims on appeal.

Table 2-2 of the Griffiths *et al.* textbook illustrates standard terminology and concepts involved in an example of simple Mendelian inheritance: that of the inheritance of color of flowers of varieties of garden pea plant with which Mendel experimented. The plant "character" which Table 2-2 addressed was flower color, which exhibited two "phenotypes" of flower color: purple and white. As noted in the table, cross pollination of plants from a pure line with purple flowers with plants from a pure line with white flowers yielded first generation offspring with plants of only purple flowers, and self pollination of plants within that first generation yielded

second generation offspring in which approximately ¾ of the plants had purple flowers and the remaining approximately ¼ of the plants had white flowers. From these and similar results, Mendel inferred that the difference in character was controlled by what is now called a "gene" for flower color that had two variant forms, now termed "alleles," one of which allele was dominant over the other allele, termed a "recessive" allele. The allele of the flower-color gene which caused purple flowers, denoted ***C*** in Table 2-2 of the Griffiths *et al.* textbook, was the dominant allele and the allele which caused white flowers, denoted *c* in the table, was the recessive allele.

Leaping ahead from the experimental techniques of Mendel to modern molecular-biology technology current as of the publication of the Griffiths *et al.* textbook (and thus evidently part of the technological background which would have been understood by persons of ordinary skill in the art as of the effective filing date of the subject patent application), it is explained at page 25 of the textbook, left-hand column, line 31 to line 7 of the right-hand-side column that:

> [w]hen alleles like ***A*** and ***a*** are examined at the DNA level using modern technology, it is generally found that they are identical for most of their sequence, and differ only at one or a few nucleotides out of the thousands that make up the gene. Therefore, we see that the alleles are truly different versions of the same gene.

Returning to the principles of Mendelian genetic analysis, as explained in the Griffiths *et al.* textbook at page 25, left-hand-side column, lines 9 to 30 and with reference to Figure 2-6 on page 24, any individual pea plant would always have <u>two copies</u> of the flower-color gene, one copy derived from each parent gamete. (Organisms such as the pea plants studied by Mendel whose genetic material carry two copies of each gene are called "diploid." An organism whose genetic material carries more than two copies of each gene is conventionally called "polyploid," e.g. a tetraploid plant such as corn has genetic material which carries four gene copies. An organism whose genetic material carries only a single copy of each gene is called "monoploid."). The two copies of the pea-plant flower-color gene, depending on random inheritance from parental gametes, would consist of one of the allele pairs denoted ***CC***, ***Cc***, or ***cc***. Regarding terminology, it is noted in the Griffiths *et al.* textbook at page 24, right-hand-side column, lines

17 through 32, that the designation of an allele pair defines a "genotype" and that different allele pairs constitute different genotypes – even in cases in which, by virtue of the presence of an allele which is dominant, the different genotypes give rise to the same phenotype. Table 2-2 indicates that the two different alleles ***C*** and ***c*** of the flower-color gene distributed among the two copies of the flower-color gene produced three different possible genotypes which determined the color of the flowers of the plants: the genotype ***CC***, identified as "homozygous dominant," and the genotype ***Cc***, identified as "heterozygous;" both of which gave rise to plants with purple flowers; and the genotype ***cc***, identified as "homozygous recessive," which gave rise to plants with white flowers. It should be noted that in certain cases more than two possible alleles may be observed with respect to a single locus across a population of biological organisms.

As demonstrated below, persons of ordinary skill in the art would have recognized that the subject patent application used the terms "genotype" and "allele" in an art-recognized conventional sense in which, in a typical diploid organism, a genotype was defined with respect to a pair of genes (respectively derived from two gamete cells; in a cross-fertilization situation, one gamete cell from each of two parents of the organism). Each gene of that gene pair could be one of at least two different alleles, so that, for such an organism, a genotype as the term would have been understood to have been used in the subject patent application would have to have been specified by either a pair of identical alleles (homozygous case) or by a pair of different alleles (heterozygous case) – entirely consistent with the principles of Mendelian genetic analysis discussed above.

In analyzing the art cited against the claims of the subject application in the Office Action on appeal and, in particular, in evaluating the various hypothetical combinations of art citations proposed in the Office Action, it is important to keep in mind that persons of ordinary skill in the art recognized fundamental distinctions between the subfields of *genetic analysis of individuals* and *population genetics* in the field of genetic analysis, as is evident from the following passage from the Griffiths *et al.* textbook cited above:

> Mendel's investigations of heredity — indeed all the interest in heredity in the nineteenth century — arose from two related problems: how to breed improved crops and how to understand the nature and origin of species. What is common to these problems (and differentiates them from the problems of transmission and gene action) is that they are concerned with *populations* rather than with *individuals*. Studies of gene replication, protein synthesis, development, and chromosome movement focus on processes that go on within cells of individual organisms. But transformation of a species, either in the natural course of evolution or by deliberate intervention of human beings, is a change in the properties of a collectivity — of an entire population or a set of populations. [Italics in original.]

Griffiths *et al.* at page 738, column 1, lines 1 through 15; included in the accompanying evidence appendix. In general, methods of analysis employed in the field of population genetics are not directly applicable to the field of genetic analysis of individuals.

As discussed in detail below, a hypothetical combination of art citations was proposed in the Office Action on appeal with comments regarding asserted motivation of the combination which made reference to the concepts of "Hardy-Weinberg equilibrium" and a "Hardy-Weinberg test." It is thus important to understand what persons of ordinary skill in the art as of the effective filing date of the subject application would have understood these two Hardy-Weinberg concepts to involve.

As may be seen from pages 750 through 752 of the Griffiths *et al.* textbook included in the accompanying evidence appendix, Hardy-Weinberg equilibrium refers to the statistical distribution of genotypes associated with a given genetic locus over a population of individuals who are the offspring from a population of parents who respectively selected mates without regard to the genotype of the mate with respect to the locus. Thus Hardy-Weinberg equilibrium is a concept employed in the field of population genetics, as opposed to being of use in connection with the genetic analysis of an individual. Hardy-Weinberg equilibrium is defined such that, in the case of a dialleic genetic locus having a first allele denoted A with a frequency of p over the population of the parent generation and a second allele denoted a having a frequency $q = 1\text{-}p$ over the parent-generation population, the frequencies of the three resulting

genotypes over the population of the offspring generation under Hardy-Weinberg equilibrium would be as follows: the *AA* homozygotes would have a frequency of p^2; the *Aa* heterozygotes would have a frequency of *2pq*; and the *aa* homozygotes would have a frequency of q^2. If, in contrast to the random selection of mates vis-à-vis the genotypes associated with the genetic locus under discussion posited for Hardy-Weinberg equilibrium, members of the parent generation tended to select mates on the basis of the selected mate's exhibiting a trait like their own which was genetically determined by the genotypes associated with the locus, the frequencies of the genotypes over the population of the offspring generation would tend to deviate from the frequencies specified by Hardy-Weinberg equilibrium.

Regarding the concept of Hardy-Weinberg test, it is submitted that persons of ordinary skill in the art would have understood the expression "Hardy-Weinberg test" to refer to a test for Hardy-Weinberg equilibrium in a population of interest involving ascertaining either the frequencies or aggregate numbers of genotypes and constituent alleles exhibited in a sample of individuals randomly selected from the population and comparing in some way the observed frequencies of the various genotypes exhibited in the sample to the corresponding "expected" genotype frequencies computed from the frequencies of alleles exhibited in the sample using the set of algebraic relations which define Hardy-Weinberg equilibrium to make a judgment about the extent to which the overall population of interest exhibits or departs from Hardy-Weinberg equilibrium. See Table 25-11 on page 752 of Griffiths *et al.* textbook and the discussion regarding Table 25-11 on page 751 of the textbook. A test for Hardy-Weinberg equilibrium would ordinarily take as original input data, data enumerating aggregate numbers of genotypes and constituent alleles exhibited in a population sample based on previously obtained genotype identifications for members of the sample, and would not provide any output information which could be used by itself to determine the particular genotype of any individual from the population.

VII.*a.2*) The Kimpton *et al.* Publication in View of the Ledwina *et al.* Publication as Assertedly Motivated in View of the Jeanpierre Publication

VII.*a.2.i*) Summary of Office Action on Appeal with Respect to Final Rejection Under 35 U.S.C. § 103(a) of Claims as Unpatentable over Kimpton *et al.* Publication in View of the Ledwina *et al.* Publication as Assertedly Motivated in View of the Jeanpierre Publication

In the Office Action of 12 October 2005 on appeal, pending claims 75, 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, 95, 96 through 98 inclusive, 102, 106 through 109 inclusive, and 112 through 115 inclusive were finally rejected under 35 U.S.C. §103(a) as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication.

It was asserted in the Office Action on appeal that the Kimpton *et al.* publication disclosed at page 14 a method of determining the genotype at a locus within genetic material obtained by PCR amplification. With citations to page 14, columns 1 through 3 of the Kimpton *et al.* publication, it was asserted that a method of the publication included reacting material at the locus to produce a first reaction value. It was asserted further, with a citation to pages 14 through 16, that the publication disclosed forming a data set including the first reaction value by assembling reaction value data points for samples with each reaction value data point assertedly corresponding to a respective one of the samples and including at least one reaction value. It was asserted that the data points represented by each of the separate peaks in Figure 1 of the Kimpton *et al.* publication represented a different sample and were assembled in Figure 2. With a reference to pages 16 and 17 of the Kimpton *et al.* publication, it was asserted that the publication disclosed determining the genotype and a confidence score for each reaction value data point assertedly to determine the genotype and confidence score at the genetic locus for each sample. It was asserted that Table 2 on page 17 of the publication provided for each reaction point the genotype and a standard deviation based on the data obtained from a "step d," which "step d" was not further identified in the Office Action.

With regard to claim 78, it was asserted in the Office Action on appeal that the Kimpton *et al.* publication disclosed reacting the material at multiple loci at page 14, Table 1.

With regard to claims 80 through 82, 114, and 115, it was asserted that the Kimpton *et al.* publication on page 17 "expressly considered" multiple alleles in probability distributions; however, such probability distributions were not identified further in the Office Action. Table 2 was cited with the assertion that the Table disclosed that the method of the publication was applicable to any number of alleles. It was asserted in the Office Action on appeal that the Kimpton *et al.* publication expressly disclosed the use of multiple data points derived from multiple runs of an automated apparatus including multiple data sets in the method and apparatus disclosed at page 16 and in Figure 2.

Regarding claims 85, 97, 98, 108, and 109, it was asserted in the Office Action on appeal that the Kimpton *et al.* publication disclosed determination of a confidence score. Pages 16 and 17 of the Kimpton *et al.* publication were referenced in connection with the assertion concerning disclosure of a confidence score determination in the publication.

With the comment that the Kimpton *et al.* publication disclosed the use of a Hardy-Weinberg "test," it was admitted in the Office Action on appeal that the publication did not disclose establishing a distribution set of probability distributions and did not disclose applying the reaction value of the distributions to determine a measure of the conditional probability of each genotype of interest at the locus. It does not appear possible to square the admission in the Office Action on appeal that the Kimpton *et al.* publication did not disclose establishing a distribution set of probability distributions with the assertion in the Office Action noted above with respect to claims 80 through 82, 114, and 115 that the publication somehow considered multiple alleles in probability distributions.

It was asserted in the Office Action on appeal that the Ledwina *et al.* publication disclosed a method in which genotypes could be determined in which the Hardy-Weinberg "test" was modified assertedly to include the steps of establishing a distribution set of probability

distributions and associating hypothetical values of some variable which was not identified in the Office Action with corresponding probabilities for each genotype of interest. Pages 162 and 163 of the Ledwina *et al.* publication were cited in this regard. It was asserted further that the Ledwina *et al.* publication disclosed the step of applying the unidentified first value to each pertinent probability distribution to determine a measure of conditional probabilities of each genotype of interest, citing in particular pages 162 and 163 of the Ledwina *et al.* publication. It was asserted further in the Office Action on 12 October 2005 that, with respect to claims 76 and 79, the Ledwina *et al.* publication, in referring on page 162 to "common probability distribution of (**T**, **Z**) is multinomial with ½m(m+1) cells and with the vector of cell probabilities **g** = (g...)," assertedly disclosed a plurality of distributions which were hypothetical.

It was asserted in the Office Action on appeal that the Jeanpierre publication motivated the use of computation of unknown genotypes to analyze conditional probabilities relative to a distribution of hypothetical reaction values. Page 330 of the Jeanpierre publication was cited in this connection.

It was asserted in the Office Action on appeal that it would have been *prima-facie* obvious to one of ordinary skill in the art at the time the invention was made to modify the method of the Kimpton *et al.* publication to use a conditional probability distribution method assertedly disclosed in the Ledwina *et al.* publication, since the Kimpton *et al.* publication assertedly noted that the analysis of the publication used Hardy-Weinberg equilibria and since the Ledwina *et al.* publication disclosed a class of admissible tests for the Hardy-Weinberg equilibrium in a multiple allelic system. It was asserted that an ordinary practitioner would have been motivated to apply the asserted hypothetical distribution analysis to genotyping since the Jeanpierre publication assertedly disclosed certain gains from creating such a distribution, including avoiding a hazard of incorrectly using a simple average of conditional probabilities instead of a harmonic mean.

VII.*a.2.ii*) Claims 75, 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, and 95

Turning first to independent claim 75, as discussed below, none of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication individually disclosed certain elements of the method of independent claim 75 – for example, elements involving probability distributions associating reaction values with corresponding probabilities for each genotype of interest at a genetic locus – and moreover the three publications considered singly or in any combination would not have suggested such elements to a person of ordinary skill in the art as of the effective filing date of the subject application. Thus, not even a *prima facie* case of unpatentability for obviousness of claim 75 with respect to the three cited publications was made in the Office Action on appeal.

Independent claim 75 of the subject application is directed to a method for determining the genotype of a subject at a genetic locus within genetic material obtained from a biological sample from the subject which includes a step, among others, of establishing a distribution set of probability distributions associating hypothetical reaction values with corresponding probabilities for each genotype of interest at the locus. The method of claim 75 further includes the steps of reacting the material at the locus to obtain a first reaction value indicative of the presence of a given allele at the locus and applying the first reaction value to each pertinent probability distribution to determine a measure of a conditional probability of each genotype of interest at the locus. The method of claim 75 also includes a step of determining the genotype based on data from the step of applying the first reaction value to each pertinent probability distribution.

As noted in the preceding subsection, it was admitted in the fifth paragraph on page 5 of the Office Action on appeal that the Kimpton *et al.* publication did not disclose establishing a distribution set of probability distributions. The Kimpton *et al.* publication did not therefore disclose determining a measure of a conditional probability of each genotype of interest at a locus by applying reaction values to the distribution, as also admitted in the fifth paragraph on

page 5 of the Office Action. The attorneys for the applicants fully agree that the Kimpton *et al.* publication did not disclose the establishment of a distribution set of probability distributions, and submit in addition that the Ledwina *et al.* publication does not cure the admitted infirmities of the Kimpton *et al.* publication as a reference against claim 75, for the reasons discussed below.

The Kimpton *et al.* publication disclosed automated DNA profiling, based on detection of amplified tri-, tetra-, and pentanucleotide short-tandem-repeat ("STR") alleles from selected STR loci by electrophoresis on denaturing polyacrylamide sequencing gels using automated fluorescence-based technology. According to the abstract of the publication, the system of the publication used an internal size standard in each sample to permit the short-tandem-repeat products amplified by a polymerase chain reaction ("PCR") to be sized automatically. In the method of the Kimpton *et al.* publication, three multiplex short-tandem-repeat systems involving a total of fourteen different loci were used, with different fluorescent markers used for loci which had overlapping allele size ranges.

According to page 13, column 3, lines 13 through 19 of the Kimpton *et al.* publication, the ability to resolve PCR products differing in size by just one base allowed precise allele designation for short-tandem-repeat loci. However, the publication disclosed that, with respect to amplified dinucleotide short-tandem-repeat products, the amplification process gave rise to artifactual "stutter" bands which complicated allele designation. According to the Kimpton *et al.* publication at page 13, column 3, lines 13 through 29, the complication of artifactual stutter bands could be avoided by using short-tandem-repeat loci with tri- and tetrameric repeats which had wider allele spacings than short-tandem-repeat loci with dinucleotide repeats. In the words of the publication:

> [T]he ability to resolve PCR products differing in size by just 1 base on polyacrylamide gels allows precise allele designation, thus eliminating the need for continuous allele distribution models currently employed with VNTR [variable number tandem repeat] systems.

> Analysis of dinucleotide STRs has revealed enzyme slippage during amplification, resulting in artifactual "stutter" bands. This makes unambiguous allele designation difficult. However, tri- and tetrameric repeats, which have a wider allele spacing, appear to be significantly less prone to slippage and are therefore more suitable for individual identification. [Footnotes omitted.]

It was disclosed at page 16, column 1, line 51 through column 2, line 15 of the Kimpton *et al.* publication that for twelve of the fourteen short-tandem-repeat loci, the maximum band-size range was sufficiently small relative to the minimum repeat-unit size to permit unambiguous allele designation. For the remaining two loci, according to page 16, column 2, line 16 through column 3, line 14 and page 19, column 3, line 56 through page 20, column 1, line 4 of the publication, variability between polyacrylamide gels did not allow reliable allele designation even though differences between allele bands were readily resolvable on the gels. Nonetheless, according to the publication, allele designation could be accomplished for the two loci in question by running an allelic-ladder control on each gel for the two loci. Thus even for the two short-tandem-repeat loci for which variability between gels did not allow reliable allele designation, the Kimpton *et al.* publication did not suggest that any alternative method was needed to resolve the allele bands for the two loci, but disclosed that satisfactory allele designation could be accomplished with the computer-generated band-sizing technology used for the other loci by direct comparison to allelic-ladder controls run on each gel.

The quotation above from the Kimpton *et al.* publication regarding the choice of short-tandem-repeat loci with tri- and tetrameric repeats is significant with respect to the rejections relying on the publication in the Office Action of 12 October 2005 in that the quotation counters any notion that a person of ordinary skill using the analytical method disclosed in the publication would have been motivated to modify the method to determine allele sizes by some other procedure. In the analytical method disclosed in the Kimpton *et al.* publication, short-tandem-repeat loci were deliberately selected in order to permit precise, unambiguous allele designation using polyacrylamide gels with automated fluorescence-based technology and with a view to eliminating any need for continuous allele distribution models. A precise, unambiguous method to designate alleles such as disclosed in the Kimpton *et al.* publication provided all the resolution

and accuracy needed in the automated DNA profiling system of the publication, as a person of ordinary skill in the art would have recognized. A person of ordinary skill purporting to use the analytical method disclosed in the Kimpton *et al.* publication who looked for an alternative method to determine allele sizes involving some sort of continuous allele distribution model would have been going against the plain disclosure of the publication read as a whole.

The Court of Appeals for the Federal Circuit has held that in analyzing the differences between a claimed invention and the disclosure of a prior-art reference pursuant to 35 U.S.C. § 103, the prior-art disclosure must be considered in its entirety, including portions that argue against obviousness. *Bausch & Lomb v. Barnes-Hind/Hydrocurve*, 230 USPQ 416, 420 (Fed. Cir. 1986). The court in *Bausch & Lomb* quoted with approval an earlier decision by the Court of Customs and Patent Appeals, which held:

> It is impermissible within the framework of section 103 to pick and choose from any one reference only so much of it as will support a given position to the exclusion of other parts necessary to the full appreciation of what such reference fairly suggests to one skilled in the art.

In re Wesslau, 147 USPQ 391, 393 (CCPA 1965). To fail to consider a prior-art disclosure in its entirety, including portions that argue against obviousness, constitutes improper hindsight analysis according to the court in *Bausch & Lomb*.

With respect to the Kimpton *et al.* publication, it is submitted that the publication did not merely fail to disclose establishing a distribution set of probability distributions associating hypothetical reaction values with corresponding probabilities for each genotype of interest at a locus within genetic material in connection with a method involving determining the genotype of a subject at the locus using a reaction value indicative of a given allele at the locus, but, read as a whole as required under applicable appellate-court precedent noted above, would have affirmatively led persons of ordinary skill in the art away from use of probability distributions in connection with such a genotype-determination method. To ignore the disclosure of the Kimpton *et al.* publication that short-tandem-repeat loci for the genotype-determination method

of the publication be particularly selected in order to permit unambiguous allele designation using polyacrylamide gels and with a view to eliminating any need for continuous allele distribution models would be, it is submitted, to engage in impermissible hindsight analysis.

Concerning the technique for automatically sizing the short-tandem-repeat products in the procedure of the Kimpton *et al.* publication, the publication disclosed at page 16, column 1, lines 9 through 18 that amplification products of the short-tandem-repeat loci were tagged by the attachment of a fluorescent dye molecule to one of each pair of the locus-specific amplification primers. Amplification products from each of the three multiplex amplification-reaction systems, each together with a dye-labeled internal lane standard, were respectively electrophoresed on a polyacrylamide denaturing sequencing gel in an automated DNA sequencer. See page 15, column 1, lines 17 through 25 of the publication. During electrophoresis on the denaturing polyacrylamide gels, amplified products were detected by laser scanning. According to column 1, lines 17 through 34 of page 15 of the Kimpton *et al.* publication, fragment sizes after electrophoresis on the automated DNA sequencer were determined using software employing a method of second order regression to establish a curve of best fit for the internal standard in each lane. According to column 3, lines 42 through 46 of page 19 of the publication, the software sized PCR products automatically against the internal ladder standard. Other than the reference to second order regression, internal operation of the software for determining fragment sizes against an internal lane standard does not appear to be described in the Kimpton *et al.* publication.

As noted above, the attorneys for the applicants agree with the assessment in the fifth paragraph on page 5 of the Office Action of 12 October 2005 that the Kimpton *et al.* publication did not disclose establishing a distribution set of probability distributions and applying a reaction value to the distributions to determine a measure of a conditional probability of each genotype of interest at the genetic locus under investigation. Indeed, the attorneys for the applicants submit further that the Kimpton *et al.* publication not only failed to disclose establishing a distribution set of probability distributions associating hypothetical reaction values with corresponding

probabilities for each genotype of interest at a locus within genetic material in connection with a method involving determining the genotype of a subject at the locus using a reaction value indicative of a given allele at the locus, but, in disclosing that the method of the publication eliminated any need for continuous allele distribution models as discussed above, the Kimpton *et al.* publication would have affirmatively led persons of ordinary skill in the art away from use of probability distributions in connection with such a genotype-determination method.

Independent claim 75 of the subject application is directed to a method for determining the genotype of a subject at a genetic locus within genetic material obtained from a biological sample from the subject which includes a step, among others, of establishing a distribution set of probability distributions associating hypothetical reaction values with corresponding probabilities for each genotype of interest at the locus. The method of claim 75 further includes a step of applying a first reaction value indicative of the presence of a given allele at the locus to each pertinent probability distribution to determine a measure of a conditional probability of each genotype of interest at the locus. The method of claim 75 also includes a step of determining the genotype based on data from the step of applying the first reaction value to each pertinent probability distribution.

For the reasons given above, it is submitted that the Kimpton *et al.* publication as of the effective filing date of the subject application would not have disclosed the subject matter of independent claim 75 of the application, and moreover that the publication would in fact have led persons skilled in the art away from the subject matter of claim 75.

Turning now to the Ledwina *et al.* publication, it is important to keep in mind that the Ledwina *et al.* publication addressed a problem in <u>population</u> genetics, in contrast to the problem of determination of the genotype of an individual subject, to which claim 75 relates. As pointed out in the "Background Technology and Terminology" subsection above, persons of ordinary skill in the art recognized fundamental distinctions between the field of *genetic analysis of individuals* and the field of *population genetics*. The assertion in the sixth paragraph on page 5 of the Office Action on appeal that "Ledwina teaches a method in which genotypes can be

determined" was simply incorrect and mischaracterized the Ledwina *et al.* publication in a way which suggested that the publication might have been directly relevant to the method of claim 75 of determining the genotype of a subject, when in fact the publication was not.

The problem in population genetics addressed by the Ledwina *et al.* publication concerned formulating mathematically a statistically-rigorous test for accepting or rejecting the hypothesis that a particular population of individuals exhibited a certain set of algebraic relations between, on the one hand, the frequencies of genotypes over the population with respect to a genetic locus and, on the other hand, the frequencies of the component alleles of those genotypes, which set of relations defined the population-genetics state known as "Hardy-Weinberg equilibrium" discussed in the "Background Technology and Terminology" subsection above. As discussed in more detail below, the statistical test for accepting or rejecting the hypothesis that the population exhibited Hardy-Weinberg equilibrium was based on data separately enumerating the aggregate number of individuals having each different genotype in a sample of N individuals randomly selected from the population. Compiling a set of data enumerating the aggregate number of individuals from a sample having each different genotype with respect to a genetic locus requires that the genotype of each individual in the sample be determined. The particular method by which the genotype of an individual may have been determined to obtain data enumerating the aggregate number of individuals from the sample having each different genotype was entirely irrelevant to the mathematical derivation of the statistical test for Hardy-Weinberg equilibrium set out in the Ledwina *et al.* publication and no such method was specified, recommended, or suggested in the publication. In particular, no mention was made in the Ledwina *et al.* publication of obtaining reaction values in connection with any genotyping method. Consequently, as noted above, the Ledwina *et al.* publication does not cure the admitted infirmities of the Kimpton *et al.* publication as a reference against claim 75

In the course of the derivation of the statistical test for accepting or rejecting the hypothesis that a population exhibited Hardy-Weinberg equilibrium, probability distributions in several different algebraic forms were set out in the Ledwina *et al.* publication – see, for

example, equations (1), (2), and (3) of the publication on pages 162 and 163. However, each of the probability distributions in the Ledwina *et al.* publication was a function of and associated probabilities with respect to one or more random, multicomponent vectors denoted **X**, **T**, and **Z** in the publication, each of whose components, in the case of the random vectors **X** and **T**, respectively specified the aggregate number of individuals in a sample of N individuals who had a corresponding one of the possible allele-pair genotypes, denoted A_iA_j, associated with a genetic locus, or, in the case of the vector **Z**, each of whose components respectively specified the aggregate number of copies of a corresponding allele A_k represented among the genotypes A_iA_j respectively associated with the N individuals of the sample. None of the probability distributions disclosed in the Ledwina *et al.* publication concerned in any way reaction values, hypothetical or otherwise, indicative of the presence of a given allele at a locus, or concerned the identity of an individual genotype, as opposed to numbers of individuals in a sample possessing a particular genotype. More particularly, none of the probability distributions disclosed in the Ledwina *et al.* publication associated a reaction value indicative of the presence of a given allele at a locus with a probability for a genotype, as called for in claim 75 of the subject application.

An imaginary thought experiment makes it clear that the probability distributions of the Ledwina *et al.* publication could not in general have been used to determine the genotype of an individual subject, contrary to the assertions in the Office Action on appeal. In particular, the aggregate numbers specified by the components of the random vectors **X**, **T**, and **Z** of the various probability distributions of the Ledwina *et al.* publication would have remained invariant had any pair of individuals in the sample having different genotypes somehow magically swapped genotypes with one another with respect to some locus, and consequently any probabilities from probability distributions that were a function of such random vectors would have necessarily remained unchanged under such a genotype swap, demonstrating that such probability distributions were useless for determining the respective genotypes of individual subjects in samples with more than one genotype.

In contrast, as noted above, claim 75 of the subject application is directed to a method for determining the genotype of a subject which includes the steps, among others, of reacting genetic material obtained from a biological sample from the subject at a locus to produce a reaction value indicative of the presence of a given allele at the locus and applying the reaction value to each pertinent probability distribution in a distribution set of probability distributions associating hypothetical reaction values with corresponding probabilities for each genotype of interest at the locus to determine a measure of a conditional probability of each genotype of interest at the locus.

Applying our imaginary thought experiment discussed above to the method of claim 75, swapping the genotypes of two individuals prior to carrying out the method with respect to each individual would result in the respective reaction values obtained from the two individuals also being swapped, so that the measures of conditional probability for each genotype determined for each individual subject would be particular to that individual subject's genotype and could be used as claimed to determine the subject's genotype, unlike the swap-invariant probabilities from the probability distributions disclosed in the Ledwina *et al.* publication discussed above.

We turn now to the hypothetical combination of the Kimpton *et al.* publication with the Ledwina *et al.* publication proposed in the Office Action of 12 October 2005. In connection with the rejection in the Office Action under 35 U.S.C. § 103(a) of claims 75, 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, and 95, the concepts of Hardy-Weinberg equilibria and a Hardy-Weinberg test were referred to in connection with proposing the hypothetical combination of the Kimpton *et al.* publication with the Ledwina *et al.* publication.

In particular, it was asserted generally on pages 5 and 7 of the Office Action on appeal that the Kimpton *et al.* publication disclosed a method of determining the genotype at a locus within genetic material which used Hardy-Weinberg equilibria. Regarding Hardy-Weinberg equilibrium, the Kimpton *et al.* publication disclosed at page 17, left-hand-side column, lines 16 through 22 and middle column, lines 4 through 10 that allele frequencies for each of the fourteen genetic loci under investigation were determined from a minimum of 50 randomly selected

individuals from each of three different populations: Caucasian, Afro-Caribbean, and Asian. Each of these locus/population data sets was tested for Hardy-Weinberg equilibrium using a "log likelihood-G test." Of the 3x14 = 42 locus/population data sets tested, only one data set was found to deviate from Hardy-Weinberg equilibrium. The Kimpton *et al.* publication did not discuss any consequences of finding Hardy-Weinberg equilibrium or of finding a deviation from such equilibrium with respect to the locus/population data sets, other than to note at page 20, right-hand-side column, lines 23 through 35 that the one case of deviation from Hardy-Weinberg equilibrium might have been due to sampling error. The Kimpton *et al.* publication did not disclose or suggest using the test for Hardy-Weinberg equilibrium in connection with determination of the genotype of any individual. Indeed, as noted in the "Background Technology and Terminology" subsection above, any test for Hardy-Weinberg equilibrium takes as original input data genotype-enumeration data based on genotype identifications which would have to have been made previously, as, it is submitted, persons of ordinary skill in the art would have appreciated as of the effective filing date of the subject application.

In the Office Action on appeal in the sixth paragraph of page 5, it was asserted that the Ledwina *et al.* publication disclosed a method by which genotypes could be determined which somehow involved a "modified" Hardy-Weinberg test. How and for what purpose the Hardy-Weinberg test was assertedly modified in the Ledwina *et al.* publication was unclear from the Office Action. It is respectfully submitted that, contrary to the assertion in the sixth paragraph of page 5 of the Office Action on appeal, the Ledwina *et al.* publication in no way disclosed or suggested any method by which genotypes could be determined. Rather, the Ledwina *et al.* publication concerned a class of statistical tests applicable to genotype-enumeration data from a sample of N individuals whose genotypes had been previously determined to provide a basis for accepting or rejecting the hypothesis that the frequencies of genotypes over the population from which the sample of individuals had been taken exhibited Hardy-Weinberg equilibrium. More particularly, the Ledwina *et al.* publication purported to derive criteria to determine if a given statistical test for accepting or rejecting the hypothesis of Hardy-Weinberg equilibrium among the genotypes of a population based on genotype-enumeration data from a sample of N

individuals drawn from that population had statistical justification of a certain sort, which the publication referred to as being an "admissible" test. The particular test for which the greatest detail was provided in the Ledwina *et al.* publication was a so-called chi-square X^2 goodness-of-fit test. No specific mention was made in the Ledwina *et al.* publication of the log likelihood-G test used in the Kimpton *et al.* publication to test for Hardy-Weinberg equilibrium in various locus/population data sets.

In the Office Action on appeal at page 5, sixth paragraph, it was asserted, with reference to an expression for a "conditional distribution of **T** given **Z** = **z**" set out in lines 7 through 10 of page 163 of the Ledwina *et al.* publication, that the publication disclosed establishing a distribution set of probability distributions, associating hypothetical values of a sort not specified in the Office Action with corresponding probabilities for each genotype of interest, and applying a first such unspecified value to each pertinent probability distribution assertedly to determine a measure of conditional probability of each genotype of interest. A close look at the conditional distribution of **T** given **Z** = **z** disclosed in the Ledwina *et al.* publication will establish that, contrary to the assertion in the Office Action on appeal, the Ledwina *et al.* publication did not disclose or suggest determining a measure of the conditional probability of each genotype of interest with respect to a locus.

As defined in the first sentence of section 1 on page 161 of the the Ledwina *et al.* publication and the respective first sentences of the first two paragraphs of section 2, the statistic **T** in the conditional distribution of **T** given **Z** = **z** set out in lines 7 through 10 of page 163 of the publication was a random vector whose components specified, with respect to a particular genetic locus having a number m of alleles denoted A_i, the respective aggregate numbers of individuals in a sample of N individuals who had, for a first m-1 of the alleles, homozygous genotypes A_iA_i or heterozygous genotypes A_iA_j, $i \neq j$, irrespective of whichever of the alleles A_i and A_j came from the mother or the father of the individual. The statistic **T** thus could be regarded as an incomplete genotype-enumeration statistic; incomplete in the sense that only m-1 of the m alleles were accounted for directly. The statistic **Z** was defined to be a random vector

whose components specified, with respect to the same genetic locus, the respective aggregate numbers of individuals in the sample of N individuals who had, for the first m-1 of the alleles, the allele A_i, with homozygous individuals having two copies of the allele A_i being counted twice. The statistic **Z** thus could be regarded as an incomplete allele-enumeration statistic. Consequently, contrary to the assertion in the Office Action, the expression for the conditional distribution of **T** given **Z** = **z** derived in the Ledwina *et al.* publication associated hypothetical vector values of the incomplete genotype-enumeration statistic **T**, which had components which specified the respective aggregate numbers of individuals in a sample of N individuals who had genotypes A_iA_j, $1 \leq i,j \leq m$-1, irrespective of the parental source of the alleles A_i and A_j, with corresponding probabilities for each particular vector value **z** of the incomplete allele-enumeration statistic **Z**, which had components which specified the respective aggregate numbers of individuals in the sample of N individuals who had alleles A_i, $1 \leq i \leq m$-1, counting as twice homozygous individuals who had two copies of the allele A_i. Significantly, for any particular sample of N individuals, the vector components of both the incomplete genotype-enumeration statistic **T** = **t** and the incomplete allele-enumeration statistic **Z** = **z** were formed in the Ledwina *et al.* publication from previously obtained complete genotype-enumeration data $\{x_{ij}\}$, $1 \leq i,j \leq m$, for the sample, as may be seen with respect to equations (1) and (2) on page 162 of the publication. Particular methods by which the genotypes could be identified to obtain the complete genotype-enumeration data $\{x_{ij}\}$ were completely irrelevant to the derivations of the Ledwina *et al.* publication and none was specified, recommended, or suggested.

The expression in the Ledwina *et al.* publication referred to in the Office Action on appeal for the conditional distribution of the incomplete genotype-enumeration statistic **T** given a particular value **z** of the incomplete allele-enumeration statistic **Z** for a sample of N individuals set out in lines 7 through 10 of page 163 of the publication applied to such a sample whether the sample was taken from a population whose genotypic frequencies were in Hardy-Weinberg equilibrium or was taken from a population whose genotypic frequencies deviated from Hardy-Weinberg equilibrium to a greater or lesser extent. In lines 11 through 18 of page 163 of the publication, the general expression of lines 7 through 10 was reduced to the special case of an

expression, denoted $P_{\mathbf{T}|\mathbf{z},H}(\mathbf{t})$ for $\mathbf{T} = \mathbf{t}$, for the conditional distribution of **T** given $\mathbf{Z} = \mathbf{z}$ for a sample of N individuals in which the genotypic frequencies of the population from which the sample of individuals was taken were hypothesized to be in Hardy-Weinberg equilibrium. The expression $P_{\mathbf{T}|\mathbf{z},H}(\mathbf{t})$ for the conditional distribution of the incomplete genotype-enumeration statistic **T** given a particular value **z** of the incomplete allele-enumeration statistic **Z** for a sample of N individuals to be tested and further given the hypothesis that the genotypic frequencies of the population from which the sample was taken were in Hardy-Weinberg equilibrium was used in the Ledwina *et al.* publication purportedly to derive a particular chi-square X^2 goodness-of-fit test which involved dividing all possible vector values **t** of the incomplete genotype-enumeration statistic **T** into two mutually exclusive classes: a hypothesis-acceptance class $\{\mathbf{t}: X^2(\mathbf{t},\mathbf{z}) < c_\alpha\}$ and a hypothesis-rejection class $\{\mathbf{t}: X^2(\mathbf{t},\mathbf{z}) \geq c_\alpha\}$, such that the overall probability that any vector value **t** possible for a sample of N individuals for which the vector value **z** was as determined for the particular sample to be tested fell in the hypothesis-rejection class and the hypothesis that the genotypic frequencies of the population from which the sample was taken were in Hardy-Weinberg equilibrium was true would be less than a selected "significance level" α. In particular, as noted on page 164, lines 4 through 7 of the Ledwina *et al.* publication, the constant c_α used in defining the hypothesis-acceptance and hypothesis-rejection classes for the chi-square X^2 test of the publication was determined by a procedure which involved summing the conditional distribution $P_{\mathbf{T}|\mathbf{z},H}(\mathbf{s}_i)$ evaluated at certain vector values $\{\mathbf{s}_i\}$ of the incomplete genotype-enumeration statistic **T** possible for a sample of N individuals with respect to the m-allele genetic locus given the particular value **z** of the incomplete allele-enumeration statistic **Z** as determined for the particular sample to be tested. The vector values $\{\mathbf{s}_i\}$ of the statistic **T** for which the conditional distribution $P_{\mathbf{T}|\mathbf{z},H}(\mathbf{s}_i)$ was summed to determine the constant c_α were those values $\{\mathbf{s}_i\}$ having corresponding chi-square statistics $\{X^2(\mathbf{s}_i,\mathbf{z})\}$ evaluated at such values which were the greatest in magnitude relative to the chi-square statistics evaluated at all other vector values of the statistic **T** possible for a sample of N individuals given the particular value **z**.

From the discussion set forth the in the preceding paragraphs, it is evident that the assertion in the Office Action of 12 October 2005 that the Ledwina *et al.* publication disclosed

establishing a distribution set of probability distributions, associating hypothetical values of a sort not specified in the Office Action with corresponding probabilities for each genotype of interest, and applying a first such unspecified value to each pertinent probability distribution assertedly to determine a measure of conditional probability of each genotype of interest generally mischaracterized the disclosure of the publication, since, as persons of ordinary skill in the art as of the effective filing date of the subject application would have appreciated, the conditional probability distributions of the Ledwina *et al.* publication associated hypothetical vector values of an incomplete genotype-enumeration statistic **T**, which had components which specified the respective aggregate numbers of individuals in a sample of N individuals who, with respect to a particular genetic locus having a number m of alleles denoted A_i, had genotypes A_iA_j, $1 \leq i,j \leq m$-1, irrespective of the parental source of the alleles A_i and A_j, with corresponding probabilities for each particular vector value **z** of an incomplete allele-enumeration statistic **Z**, which had components which specified the respective aggregate numbers of individuals in the sample of N individuals who had alleles A_i, $1 \leq i \leq m$-1. Moreover, the conditional probability distributions of the Ledwina *et al.* publication were used in connection with the mathematical derivation of a particular test for the presence of a state of Hardy-Weinberg equilibrium in a population based on genotype-enumeration data obtained for a random sample of individuals drawn from the population. It is submitted that the Ledwina *et al.* publication neither disclosed or suggested using any test for Hardy-Weinberg equilibrium in connection with determination of the genotype of any individual, nor disclosed or suggested using any conditional distribution of **T** given **Z** = **z** for a sample of N individuals, or any other conditional probability distribution, in connection with determination of the genotype of an individual.

It was asserted on page 6 of the Office Action on appeal that it would have been obvious to one of ordinary skill in the art to modify the genotyping method of the Kimpton *et al.* publication to use a conditional probability method assertedly disclosed in the Ledwina *et al.* publication since both publications used Hardy-Weinberg equilibrium analysis in some way. However, the observation that both publications used some form of Hardy-Weinberg equilibrium analysis is a red herring with respect to the assertion that it would have been obvious to modify

the genotyping method of the Kimpton *et al.* publication somehow in view of the disclosure of the Ledwina *et al.* publication, since, as pointed out above, Hardy-Weinberg equilibrium analysis applied only to population genetics, in contrast to the determination of the genotype of an individual subject to which the method of claim 75 is directed. Hardy-Weinberg equilibrium analysis was applied in the Kimpton *et al.* publication to genotype/allele enumeration data from samples of at least 50 individuals randomly selected from each of three different populations: Caucasian, Afro-Caribbean, and Asian. As pointed out above, neither the Kimpton *et al.* publication nor the Ledwina *et al.* publication disclosed or suggested using a test for Hardy-Weinberg equilibrium in connection with the determination of the genotype of an individual subject. As persons of ordinary skill in the art would have appreciated from the Kimpton *et al.* and the Ledwina *et al.* publications, any test for Hardy-Weinberg equilibrium takes as input genotype-enumeration data for a population sample based on genotype identifications which would have been made independently of the test for Hardy-Weinberg equilibrium by some suitable genotyping method, and would not provide by itself any output which could be used to determine the particular genotype of any individual selected from the population.

To the extent it might have occurred to a person of ordinary skill in the art to combine the disclosures of the Kimpton *et al.* publication and Ledwina *et al.* publication, we submit such a person would either have evaluated the log likelihood-G test reportedly used in the Kimpton *et al.* publication to test locus/population data sets for Hardy-Weinberg equilibria according to the admissibility criteria disclosed in the Ledwina *et al.* publication or would have substituted the particular chi-square X^2 goodness-of-fit test disclosed in the Ledwina *et al.* publication for the log likelihood-G test used in the Kimpton *et al.* publication. It is submitted that neither such hypothetical combination would have met the limitations of independent claim 75 of the subject application, nor would have suggested the claimed subject matter to a person of ordinary skill in the art as of the effective filing date of the application, since, for example, as noted above, neither the Kimpton *et al.* publication nor Ledwina *et al.* publication disclosed a distribution set of probability distributions associating hypothetical reaction values with corresponding probabilities for each genotype of interest at a locus within genetic material in connection with a

method involving determining the genotype of a subject at the locus using a reaction value indicative of a given allele at the locus.

Turning now to the Jeanpierre publication, the Jeanpierre publication disclosed a method for deriving the probabilities of genotypes involving DNA markers linked to a mutant genetic-disease gene for a particular member of a multigenerational family who was deceased, uncooperative, or otherwise unavailable for a direct genotype determination with respect to the markers, but whose status with respect to expressing the genetic disease was known. To determine the genotype probabilities of such an unavailable family member – referred to as an "unsampled" individual in the Jeanpierre publication – the method of the publication made use of genotype assignments of relatives in the pedigree of the family member both with respect to the DNA markers and with respect to expression of the genetic disease, as well as the status of the unsampled family member with respect to having the genetic disease. See Figures 1 and 2 on pages 327 and 328 of the Jeanpierre publication.

Particular methods by which the genotypes could be assigned to relatives who were sampled in the pedigree of the unsampled family member, other than observation of the expression of the particular genetic disease in such relatives, were irrelevant to the method of the Jeanpierre publication, and none was specified, recommended, or suggested.

In the second paragraph on page 6 of the Office Action on appeal, it was asserted that the Jeanpierre publication would have motivated "the use of computation of unknown genotypes to analyze the conditional probabilities relative to a distribution of hypothetical reaction values." However, like the Ledwina *et al.* publication, the Jeanpierre publication made no mention whatsoever of reaction values, let alone of probability distributions associating reaction values, hypothetical or otherwise, with corresponding probabilities, conditional or otherwise. Moreover, the Jeanpierre publication did not disclose the probability distribution of any variable even analogous to a reaction value. It is submitted, therefore, that the Jeanpierre publication would not have provided any motivation regarding conditional probabilities relative to a distribution of hypothetical reaction values.

Regarding the conditional probabilities disclosed in the Jeanpierre publication, each specific conditional probability disclosed involved the genotypes of two different people in a pedigree of related individuals – a very different sort of conditional probability than the conditional probability called for in claim 75 and in no way suggestive of the probabilities of claim 75. For example, each of the various conditional probabilities specifically disclosed in the Jeanpierre publication: $P(NC \mid H_1)$, $P(NC \mid H_2)$, $P(H_1|NC)$, $P(H_2|NC)$, $P(X| H_1)$, $P(X| H_2)$, $P(H_1|X)$, and $P(H_2|X)$ which were set out on page 326 of the publication; was defined in terms of a genotype of one parent in a pedigree given a genotype of the other parent and given known genotype data from the pedigree. Thus, for example, $P(NC \mid H_1)$ in equation 1 on page 326 of the Jeanpierre publication was defined in page 325, line 33, through page 326, line 5 and page 326, line 18 to be the conditional probability that the spouse of the unsampled family member was not a carrier of the mutated gene causing a lethal X-linked disease, given the hypothesis that the unsampled family member presented a particular one of two DNA-marker alleles; specifically, the allele designated "allele 1"; and given the other known pedigree data, including the status of the unsampled individual with respect to expressing the genetic disease. To give a concrete example, $P(NC \mid H_1)$ would give the conditional probability that Mrs. Jones was not a carrier of the X-linked disease (1) given the hypothesis that her spouse Mr. Jones had allele 1 relative to a particular genetic-marker locus for the disease, and (2) given that Mr. Jones did not suffer from the disease, and (3) given that the rest of the family of Mr. and Mrs. Jones – ancestors, siblings, and offspring – had genotypes with respect to the genetic-marker locus and expression of the genetic disease as specified in a Jones-family pedigree – such as, for example, the pedigree illustrated in Figure 1 on page 327 of the Jeanpierre publication.

In the abstract of the Jeanpierre publication, it was asserted that, according to the genotype reconstruction procedure of the publication, a "compound risk" was better represented as the harmonic mean of conditional probabilities than the average of the probabilities. In the third paragraph of page 6 of the Office Action on appeal it was asserted that an ordinary practitioner would have been motivated to apply the distribution analysis of the Ledwina *et al.* publication to the genotyping method of the Kimpton *et al.* publication to avoid "hazards such as

incorrectly using the simple average of the conditional probabilities instead of the harmonic mean." It was not explained in the Office Action where the hazard of averaging conditional probabilities might have arisen in the genotyping method of the Kimpton *et al.* publication, which did not involve the use of conditional probabilities. It is submitted that an ordinary practitioner, rather than modifying the genotyping method of the Kimpton *et al.* publication to avoid a hazard that the method didn't present, would have relied on the assurance set out in the publication that the short-tandem-repeat loci selected for the method permitted precise, unambiguous allele designation using polyacrylamide gels with the automated fluorescence-based technology disclosed in the publication, and so eliminated any need for continuous allele distribution models, and would have practiced the method without the modification proposed in the Office Action on appeal.

It is submitted that the assertion in the Office Action on appeal that the Jeanpierre publication would somehow have motivated analysis of distributions of hypothetical reaction values nowhere disclosed in the publication nor in the Ledwina *et al.* publication, and taught away from by the Kimpton *et al.* publication, constituted an impermissible importation of subject matter from the disclosure of the subject application and represented an improper hindsight recreation of the invention of claim 75.

On page 9 of the Office Action on appeal, it was asserted that the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication were all concerned with genotyping of humans and with the statistical analyses necessary for proper genotyping, and that each of the publications was assertedly pertinent to the same issues and problems. However, the "high-altitude aerial overview" of the Kimpton *et al.*, Ledwina *et al.*, and Jeanpierre publications represented by the assertion in the Office Action on appeal that the three publications were all concerned with genotyping of humans and with the statistical analyses necessary for proper genotyping was seriously distorted. For example, such statistical analysis as the Kimpton *et al.* publication presented concerned the effectiveness of the DNA profiling system of the publication overall and the likelihood that two different individuals would have the

same DNA profile with respect to the fourteen short-tandem-repeat loci used for the system. The Kimpton *et al.* publication did not disclose or suggest the use of any sort of statistical analysis with respect to assigning a DNA profile to a particular individual; on the contrary, as discussed above, the fourteen short-tandem-repeat loci were deliberately selected for the DNA profiling system of the Kimpton *et al.* publication in order to permit precise, unambiguous allele designation using polyacrylamide gels with automated fluorescence-based technology and with a view to eliminating any need for continuous allele distribution models. The Ledwina *et al.* publication was in no way concerned with "statistical analyses necessary for proper genotyping." As explained above, the Ledwina *et al.* publication used a statistical analysis in addressing a problem in population genetics; namely, whether a population exhibited Hardy-Weinberg equilibrium; not in connection with the proper determination of the genotype of an individual.

Each of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication was concerned with a different and disparate sort of genetic analysis: determination of a multilocus DNA profile of an individual from a sample of genetic material obtained from the individual, in the case of the Kimpton *et al.* publication; whether the respective aggregate numbers of different genotypes previously determined for a sample of N individuals drawn from a population of interest supports in a statistically-valid sense the hypothesis that the population exhibits a Hardy-Weinberg equilibrium state, in the case of the Ledwina *et al.* publication; and evaluation of the probability of a DNA-marker genotype of an individual whose genetic material cannot be obtained for testing from the genotypes of family members in the pedigree of the individual, in the case of the Jeanpierre publication. It is submitted that in view of the fundamentally different sorts of genetic analysis disclosed in the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, it would never have occurred to a person of ordinary skill in the art as of the effective filing date of the subject application to attempt to combine these three publications as proposed in the Office Action on appeal.

Assuming for the sake of argument only that it might have occurred to a person of ordinary skill in the art as of the effective filing date of the application to attempt to combine the

respective disclosures of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, it is submitted that any resulting hypothetical combination would not have met the limitations of independent claim 75 of the subject application, since, for example, not one of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication disclosed or suggested a distribution set of probability distributions associating hypothetical reaction values with corresponding probabilities for each genotype of interest at a locus within genetic material in connection with a method involving determining the genotype of a subject at the locus using a reaction value indicative of a given allele at the locus – indeed, as pointed out above, neither the Ledwina *et al.* publication nor the Jeanpierre publication made any mention whatsoever of reaction values, the assertions in the Office Action on appeal to the contrary notwithstanding.

For the reasons set forth above, it is submitted that the Kimpton *et al.* publication considered alone or in any combination with the Ledwina *et al.* publication and/or the Jeanpierre publication would not have disclosed or suggested the method of independent claim 75 of the subject application to a person of ordinary skill in the art as of the effective filing date of the application. The rejection under 35 U.S.C. § 103(a) of independent claim 75 as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was without justification, it is submitted, and should be reversed.

Each of claims 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, and 95 is a dependent claim which respectively depends directly or indirectly on independent claim 75 and consequently incorporates the limitations of claim 75 by reference. The reasoning set forth above concerning distinctions between the Kimpton *et al.* publication considered alone or in combination with the Ledwina *et al.* publication or the Jeanpierre publication and the method of independent claim 75 therefore applies with equal force with respect to dependent claims 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, and 95.

Moreover, each of dependent claims 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, and 95 includes additional subject matter which, it is submitted, renders the method of the claim even more unobvious than the method of independent claim 75 with respect to the hypothetical combination of the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication proposed in the Office Action on appeal. For example, claim 85 of the subject application is directed to the method of claim 75 in which the step of determining the genotype of a subject at a locus based on data from the step of applying a reaction value indicative of the presence of a given allele at the locus to each pertinent probability distribution associating hypothetical reaction values with corresponding probabilities for each genotype of interest to determine a measure of a conditional probability of each genotype of interest at the locus includes the step of calculating a confidence score associated with the determination of the genotype based on data from the step of applying the reaction value to each pertinent probability distribution. Contrary to the assertion in the sixth paragraph of page 4 of the Office Action on appeal in connection with the rejection of claim 85, the Kimpton *et al.* publication did not disclose, or suggest, any determination of a confidence score, let alone a determination of a confidence score based on data from a step of applying a reaction value to each pertinent probability distribution as called for in claim 85. As pointed out above, the Kimpton *et al.* publication disclosed that the short-tandem-repeat loci for the DNA profiling system of the publication were selected in order to permit precise, unambiguous allele designation using polyacrylamide gels with automated fluorescence-based technology. The Kimpton *et al.* publication did not disclose or suggest that a confidence score be associated with the DNA profile incorporating such purportedly precise, unambiguous allele designations for an individual profiled pursuant to the method of the publication.

In sum, for the reasons set forth above, it is submitted that the final rejection under 35 U.S.C. § 103(a) of claims 75, 76, 78 through 82 inclusive, 85, 86, 91 through 93 inclusive, and 95 of the subject application as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was unwarranted and should be reversed.

VII.*a.2.iii*) Claims 96 through 98 inclusive and 102

Broadly, independent claim 96 of the subject application is directed to a method of associating with a sample of genetic material from a subject one of a predetermined plurality of genotypic classes defined with respect to a genetic locus sited in the genetic material together with a corresponding confidence measure. The method of claim 96 includes a step, among others, of carrying out one or more allele-sensitive reactions on the genetic material of the sample at the genetic locus to obtain a plurality of quantitative allele-indicative reaction values. Each allele-indicative reaction value is indicative of the likely presence or absence of a particular allele defined with respect to the genetic locus. The plurality of reaction values corresponding to the sample are assembled as a reaction-value data point. The method of claim 96 includes a further step of obtaining with respect to each of the genotypic classes corresponding reaction-value data-point conditional-probability-measure distribution information, which provides, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class. The method of claim 96 also includes the step of evaluating for each of the genotypic classes the corresponding reaction-value data-point conditional-probability-measure distribution information with respect to the allele-indicative reaction values of the reaction-value data point corresponding to the sample, to obtain for each of the corresponding genotypic classes a reaction-value data-point conditional probability measure of the reaction-value data point, given the genotypic class. The method of claim 96 further comprises the step of establishing with respect to each genotypic class a genotypic-class conditional probability measure of the genotypic class given the reaction-value data point corresponding to the sample from the reaction-value data-point conditional probability measure of the reaction-value data point given the genotypic class. The method of claim 96 also includes the step of selecting a highest-probability genotypic-class conditional probability measure of a genotypic class given the reaction-value data point having a highest probability value from the genotypic-class conditional probabilities to identify a most-likely genotypic class. Finally, the method of independent claim 96 includes the step of establishing a confidence measure for the association of the most-likely

genotypic class with the sample. The confidence measure is established from values of the genotypic-class conditional probability measures of the respective genotypic classes given the reaction-value data point corresponding to the sample.

Since independent claim 96 calls for, among other things, obtaining with respect to each of the genotypic classes corresponding conditional-probability-measure distribution information, which provides, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class and evaluating, for each of the genotypic classes, the corresponding conditional-probability-measure distribution information with respect to the allele-indicative reaction values of the reaction-value data point corresponding to the sample, to obtain for each of the corresponding genotypic classes a reaction-value data-point conditional probability measure of the reaction-value data point, given the genotypic class, the reasoning set forth in the preceding subsection with respect to distinctions between the art cited in the Office Action on appeal and independent claim 75 applies with equal force to claim 96. For conciseness, that reasoning will only be summarized below. For more detail concerning the cited art, the preceding subsection should be consulted.

As discussed in the preceding subsection, in the fifth paragraph on page 5 of the Office Action on appeal it was admitted that the Kimpton *et al.* publication did not disclose establishing a distribution set of probability distributions and applying a reaction value to the distributions to determine a measure of a conditional probability of each genotype of interest at the genetic locus under investigation. The attorneys for the applicants contend that the Kimpton *et al.* publication not only failed to disclose obtaining with respect to each of a plurality of genotypic classes corresponding conditional-probability-measure distribution information, which provides, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class, but, in disclosing that the method of the publication eliminated any need for continuous allele distribution models as discussed above, the Kimpton *et al.* publication would have affirmatively

led persons of ordinary skill in the art away from use of conditional-probability-measure distribution information in connection with a genotype-determination method such as the method of independent claim 96.

The attorneys for the applicants submit that the Ledwina *et al.* publication does not cure the admitted infirmities of the Kimpton *et al.* publication as a reference against claim 96. The problem in population genetics addressed by the Ledwina *et al.* publication concerned formulating mathematically a statistically-rigorous test for accepting or rejecting the hypothesis that a particular population of individuals exhibited the set of algebraic relations between the frequencies of genotypes over the population with respect to a genetic locus and the frequencies of the component alleles of those genotypes which defined Hardy-Weinberg equilibrium. As discussed in detail in the preceding subsection, the statistical test for accepting or rejecting the hypothesis that the population exhibited Hardy-Weinberg equilibrium of the Ledwina *et al.* publication was based on data separately enumerating the aggregate number of individuals having each different genotype in a sample of N individuals randomly selected from the population. Additionally, the derivation of the test entailed manipulating certain probability distributions which were function of and associated probabilities with respect to one or more random, multicomponent vectors, each of whose components respectively specified either the aggregate number of individuals in a sample of N individuals who had a corresponding one of the possible allele-pair genotypes A_iA_j associated with a genetic locus, or the aggregate number of copies of a corresponding allele A_k represented among the genotypes A_iA_j respectively associated with the N individuals of the sample. Data separately enumerating the aggregate number of individuals having each different genotype or allele in a sample of individuals requires that the genotype of each individual in the sample be determined. However, the particular method by which the genotype of an individual may have been determined to obtain data enumerating the aggregate number of individuals from the sample having each different genotype was entirely irrelevant to the mathematical derivation of the statistical test for Hardy-Weinberg equilibrium set out in the Ledwina *et al.* publication and no such method was specified, recommended, or suggested in the publication. In particular, no mention was made in

the Ledwina *et al.* publication of obtaining reaction values in connection with any genotyping method. Consequently, as noted above, the Ledwina *et al.* publication does not cure the admitted infirmities of the Kimpton *et al.* publication as a reference against claim 96.

Like the Ledwina *et al.* publication, the Jeanpierre publication does not cure the admitted infirmities of the Kimpton *et al.* publication as a reference against independent claim 96. The Jeanpierre publication disclosed a method for deriving the probabilities of genotypes involving DNA markers linked to a mutant genetic-disease gene for a particular member of a family who was unavailable for a direct genotype determination with respect to the markers, but whose status with respect to expressing the genetic disease was known – the so-called "unsampled" family member. To determine the genotype probabilities of such an unavailable family member, the method of the Jeanpierre publication made use of genotype assignments of relatives in the pedigree of the family member both with respect to the DNA markers and with respect to expression of the genetic disease, as well as the status of the unsampled family member with respect to having the genetic disease. Particular methods by which the genotypes could be assigned to relatives who were sampled in the pedigree of the unsampled family member, other than observation of the expression of the particular genetic disease in such relatives, were irrelevant to the method of the Jeanpierre publication, and none was specified, recommended, or suggested.

Like the Ledwina *et al.* publication, the Jeanpierre publication made no mention whatsoever of reaction values, let alone of obtaining with respect to each of a plurality of genotypic classes corresponding conditional-probability-measure distribution information, which provides, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class. The Jeanpierre publication did not disclose the probability distribution of any variable even analogous to a reaction value. The disclosure of the Jeanpierre publication would have in no way countered the disclosure of the Kimpton *et al.* publication that a desirable feature of the genotyping method of the Kimpton *et al.* publication was the elimination of any need for

continuous allele distribution models. It is submitted, therefore, that the Jeanpierre publication would not have provided any motivation regarding conditional-probability-measure distribution information providing, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of hypothetical reaction-value data points given a genotypic class, contrary to the assertion in the Office Action on appeal.

As pointed out in the preceding subsection, each of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication was concerned with a different and disparate sort of genetic analysis: determination of a multilocus DNA profile of an individual from a sample of genetic material obtained from the individual, in the case of the Kimpton *et al.* publication; whether the respective aggregate numbers of different genotypes previously determined for a sample of *N* individuals drawn from a population of interest supports in a statistically-valid sense the hypothesis that the population exhibits a Hardy-Weinberg equilibrium state, in the case of the Ledwina *et al.* publication; and evaluation of the probability of a DNA-marker genotype of an individual whose genetic material cannot be obtained for testing from the genotypes of family members in the pedigree of the individual, in the case of the Jeanpierre publication. It is submitted that in view of the fundamentally different sorts of genetic analysis disclosed in the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, it would never have occurred to a person of ordinary skill in the art as of the effective filing date of the subject application to attempt to combine these three publications as proposed in the Office Action on appeal with respect to the final rejection of claim 96.

Assuming for the sake of argument only that it might have occurred to a person of ordinary skill in the art as of the effective filing date of the application to attempt to combine the respective disclosures of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, it is submitted that any resulting hypothetical combination would not have met the limitations of independent claim 96 of the subject application, since, for example, not one of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication disclosed or suggested obtaining with respect to each of a plurality of genotypic

classes corresponding conditional-probability-measure distribution information providing over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class in connection with a method involving associating a genotypic class with a sample of genetic material from a subject – indeed, as pointed out above, neither the Ledwina *et al.* publication nor the Jeanpierre publication made any mention whatsoever of reaction values, the assertions in the Office Action on appeal to the contrary notwithstanding.

As noted above, the method of independent claim 96 includes the step of establishing a confidence measure for the association of the most-likely genotypic class with the sample. The confidence measure is established from values of the genotypic-class conditional probability measures of the respective genotypic classes given the reaction-value data point corresponding to the sample. The Kimpton *et al.* publication did not disclose, or suggest, any determination of a confidence measure for the association of a DNA profile or other genotype with an individual in accordance with the genotyping method of the publication, contrary to the assertion in the Office Action on appeal. As discussed in detail above, the Kimpton *et al.* publication disclosed that the short-tandem-repeat loci for the DNA profiling system of the publication were selected in order to permit precise, unambiguous allele designation using polyacrylamide gels with automated fluorescence-based technology. The Kimpton *et al.* publication did not disclose or suggest that a confidence measure be associated with the DNA profile incorporating such purportedly precise, unambiguous allele designations for an individual profiled pursuant to the method of the publication. The Jeanpierre publication also did not disclose or suggest the establishment of a confidence measure in connection with the reconstruction of a genotype of an unsampled family member in accordance with the procedure of the publication. The Ledwina *et al.* publication, which was in no way concerned with the assignment of a genotype to an individual, did not disclose or suggest the establishment of a confidence measure in connection with a genotype assignment.

For the reasons set forth above, it is submitted that the Kimpton *et al.* publication considered alone or in any combination with the Ledwina *et al.* publication and/or the Jeanpierre publication would not have disclosed or suggested the method of independent claim 96 of the subject application to a person of ordinary skill in the art as of the effective filing date of the application. The rejection under 35 U.S.C. § 103(a) of independent claim 96 as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was without justification, it is submitted, and should be reversed.

Each of claims 97, 98, and 102 respectively depends directly or indirectly on independent claim 96 and consequently incorporates the limitations of independent claim 96 by reference. The reasoning set forth above concerning distinctions between the Kimpton *et al.* publication considered alone or in combination with the Ledwina *et al.* publication or the Jeanpierre publication and the method of independent claim 96 therefore applies with equal force with respect to dependent claims 97, 98, and 102. Consequently, it is submitted that the Kimpton *et al.* publication considered alone or in any combination with the Ledwina *et al.* publication and/or the Jeanpierre publication would have neither disclosed nor in any way suggested the subject matter of claims 97, 98, and 102 to a person of ordinary skill in the art as of the effective filing date of the subject application. It is submitted, therefore, that the rejection under 35 U.S.C. § 103(a) of dependent claims 97, 98, and 102 of the subject application as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was without justification and should be reversed.

To summarize, for the reasons set forth above, it is submitted that the final rejection under 35 U.S.C. § 103(a) of claims 96 through 98 inclusive, and 102 of the subject application as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was unwarranted and should be reversed.

VII.*a.2.iv*) Claims 106 through 109 inclusive
and 112 through 115 inclusive

Independent claim 106 of the subject application is directed to a method of associating with a sample of genetic material from a subject (i) one of a predetermined plurality of genotypic classes defined with respect to a genetic locus sited in the genetic material and (ii) a confidence measure for the association of the genotypic class with the sample. The method of claim 106 includes the step, among others, of obtaining with respect to each of the genotypic classes corresponding reaction-value data-point conditional-probability-measure distribution information which provides, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the hypothetical reaction-value data point given the genotypic class. The method of claim 106 further includes the step of evaluating the reaction-value data-point conditional-probability-measure distribution information corresponding to a genotypic class associated with the sample with respect to the reaction-value data point corresponding to the sample to obtain a reaction-value data-point conditional probability measure of the reaction-value data point given the sample genotypic class and establishing a confidence measure for the association of the sample genotypic class with the sample using the reaction-value data-point conditional probability measure of the reaction-value data point given the sample genotypic class.

Since independent claim 106 calls for, among other things, obtaining with respect to each of the genotypic classes corresponding conditional-probability-measure distribution information, which provides, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class reasoning set forth in preceding subsections with respect to distinctions between the art cited in the Office Action on appeal and independent claims 75 and 96 regarding obtaining conditional-probability-measure distribution information with respect to each of the genotypic classes of interest applies generally in parallel to claim 106. For conciseness, that reasoning will only be summarized briefly below.

As pointed out above, each of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication was concerned with a different and disparate sort of genetic analysis. It is submitted that in view of the fundamentally different sorts of genetic analysis disclosed in the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, it would never have occurred to a person of ordinary skill in the art as of the effective filing date of the subject application to attempt to combine these three publications as proposed in the Office Action on appeal with respect to claim 106. Assuming for the sake of argument only that it might have occurred to a person of ordinary skill in the art as of the effective filing date of the application to attempt to combine the respective disclosures of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, it is submitted that any resulting hypothetical combination would not have met the limitations of independent claim 106 of the subject application, since, for example, not one of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication disclosed or suggested obtaining with respect to each of a plurality of genotypic classes corresponding conditional-probability-measure distribution information providing, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class. As pointed out above, neither the Ledwina *et al.* publication nor the Jeanpierre publication made any mention whatsoever of reaction values.

The method of independent claim 106 includes the step of establishing a confidence measure for the association of the sample genotypic class with the sample. The confidence measure is specified to be established using the reaction-value data-point conditional probability measure of the reaction-value data point corresponding to the sample given the sample genotypic class. None of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication disclosed or suggested establishing a confidence measure for the association of the sample genotypic class with the sample, as called for in claim 106. The Kimpton *et al.* publication did not disclose, or suggest, any determination of a confidence measure for the association of a DNA profile or other genotype with an individual in accordance with the DNA

profiling method of the publication, contrary to the assertion on page 4 of the Office Action on appeal. The Kimpton *et al.* publication disclosed that the short-tandem-repeat loci for the DNA profiling system of the publication were selected in order to permit precise, unambiguous allele designation using polyacrylamide gels with automated fluorescence-based technology. The Kimpton *et al.* publication did not disclose or suggest that a confidence measure be associated with the DNA profile incorporating such purportedly precise, unambiguous allele designations for an individual profiled pursuant to the method of the publication. The Jeanpierre publication likewise did not disclose or suggest the establishment of a confidence measure in connection with the reconstruction of a genotype of an unsampled family member in accordance with the procedure of the publication. The Ledwina *et al.* publication was in no way concerned with the assignment of a genotype to an individual and did not disclose or suggest the establishment of a confidence measure in connection with a genotype assignment.

Each of claims 107 through 109 inclusive and 112 through 115 inclusive respectively depends directly or indirectly on independent claim 106 and consequently incorporates the limitations of independent claim 106 by reference. The reasoning set forth above concerning distinctions between the Kimpton *et al.* publication considered alone or in combination with the Ledwina *et al.* publication or the Jeanpierre publication and the method of independent claim 106 therefore applies with equal force with respect to dependent claims 107 through 109 inclusive and 112 through 115 inclusive. Consequently, it is submitted that the Kimpton *et al.* publication considered alone or in any combination with the Ledwina *et al.* publication and/or the Jeanpierre publication would have neither disclosed nor in any way suggested the subject matter of claims 107 through 109 inclusive and 112 through 115 inclusive to a person of ordinary skill in the art as of the effective filing date of the subject application. It is submitted, therefore, that the rejection under 35 U.S.C. § 103(a) of dependent claims 107 through 109 inclusive and 112 through 115 inclusive of the subject application as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was without justification and should be reversed.

In sum, for the reasons set forth above, it is submitted that the final rejection under 35 U.S.C. § 103(a) of claims 106 through 109 inclusive and 112 through 115 inclusive of the subject application as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication was unwarranted and should be reversed.

VII.*a.3*) The Kimpton *et al.* Publication in View of the Ledwina *et al.* Publication as Assertedly Motivated in View of the Jeanpierre Publication and Further in View of the Goelet *et al.* '712 Published International Application

VII.*a.3.i*) Summary of Office Action on Appeal with Respect to Final Rejection Under 35 U.S.C. § 103(a) of Claims as Unpatentable over Kimpton *et al.* Publication in View of the Ledwina *et al.* Publication as Assertedly Motivated in View of the Jeanpierre Publication and Further in View of the Goelet *et al.* '712 Published International Application

In the Office Action of 12 October 2005, pending claims 75, 76, 78 through 82 inclusive, 85 through 87 inclusive, 91 through 98 inclusive, 100, 102, and 106 through 115 inclusive were rejected under 35 U.S.C. § 103(a) as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication, and further in view of published International Patent Application WO 92/15712 to Goelet *et al.* ("the Goelet *et al.* '712 published international application"). It was noted that the Goelet *et al.* '712 published international application disclosed genetic bit analysis methods.

Although it was asserted in the third full paragraph of page 7 of the Office Action that it would have been *prima facie* obvious to combine the method of the hypothetical combination of the Kimpton *et al.* publication in view of a publication by Clark – which Clark publication was not further identified in the Office Action – with the use of the genetic bit analysis methods of the Goelet *et al.* '712 published international application, it will be assumed in view of the second full paragraph of page 7 which referred to the Kimpton *et al.* publication in view of the

Ledwina *et al.* publication as assertedly motivated in view of the Jeanpierre publication that the reference to the Clark publication was an error and that the hypothetical combination of the Ledwina *et al.* publication as assertedly motivated in view of the Jeanpierre publication was intended instead.

In the third full paragraph of page 7 of the Office Action, a statement in the Goelet *et al.* published application on page 8, lines 27 through 33 was quoted to the effect that the method of the publication could be used to characterize nucleic acids without recourse to gel electrophoresis and could be adapted to automation. It was asserted that an ordinary practitioner would have been motivated to substitute the genetic bit analysis method for PCR amplification in order to minimize the need for gel electrophoresis and enhance the automatability of the process to speed analysis and minimize costs as assertedly motivated by the Goelet *et al.* '712 published international application.

VII.*a.3.ii*) Independent Claims 75, 96, and 106

As disclosed in the abstract, the Goelet *et al.* '712 published international application disclosed a method for determining the identity of a nucleotide base at a specific position in a nucleic acid of interest and a method for determining the presence or absence of a particular nucleotide sequence in a sample of nucleic acids. The method entailed contacting the nucleic acid of interest with an oligonucleotide primer under hybridizing conditions and treating the resulting duplex, if any, with a terminator reagent under conditions permitting base pairing of a complementary terminator present in the reagent and the occurrence of a template-dependent, primer extension reaction so as to incorporate the terminator at the 3' end of the primer. The identity of the terminator at the 3' end of the primer determined whether the hybridization occurred and the identity of the base complementary to the terminator.

As discussed in detail above, the Kimpton *et al.* publication disclosed an automated DNA profiling method which employed three multiplex groups of specially selected three to five-base pair short-tandem-repeat loci which were amplified groupwise by a polymerase chain reaction

and analyzed by denaturing polyacrylamide sequencing gels. The Ledwina *et al.* publication concerned formulating mathematically a statistically-rigorous test for accepting or rejecting the hypothesis that a particular population of individuals exhibited a certain set of algebraic relations between the frequencies of genotypes over the population with respect to a genetic locus and the frequencies of the component alleles of those genotypes, which defined the state of Hardy-Weinberg equilibrium. The Jeanpierre publication disclosed a method for deriving the probabilities of genotypes involving DNA markers linked to a mutant genetic-disease gene for a particular member of a multigenerational family who was unavailable for a direct genotype determination with respect to the markers, using of genotype assignments of relatives in the pedigree of the family member both with respect to the DNA markers and with respect to expression of the genetic disease, as well as the status of the unavailable family member with respect to expressing the genetic disease.

Like each of the Kimpton *et al.*, Ledwina *et al.*, and Jeanpierre publications discussed above, the Goelet *et al.* '712 published international application made no mention of probability distributions associating hypothetical reaction values with corresponding probabilities. Additionally, the Goelet *et al.* published application did not disclose the use of a confidence score or confidence measure with respect to determination of a genotype according to the method of the application. It is submitted that the Goelet *et al.* '712 published international application does not cure the infirmities discussed above of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication, singly or in any combination, as references against any of independent claims 75, 96 and 106 of the subject application discussed in detail in preceding subsections.

Moreover, the Goelet *et al.* published application does not connect the disparate methodologies of the Kimpton *et al.* publication, the Ledwina *et al.* publication, and the Jeanpierre publication. For the reasons discussed above, it is submitted that it would not have occurred to a person of ordinary skill in the art, as of the effective filing date of the subject application, to combine the disparate methods of the Kimpton *et al.* publication and the Ledwina

et al. publication, with or without the motivation of the Jeanpierre publication, in any combination with each other or with the Goelet *et al.* published application.

Additionally, assuming for the sake of argument only that it might have occurred to a person of ordinary skill in the art as of the effective filing date of the application to attempt to combine the respective disclosures of the Kimpton *et al.* publication, the Ledwina *et al.* publication, the Jeanpierre publication, and the Goelet *et al.* published application, any such hypothetical attempted combination would not have met the limitations of any of independent claims 75, 96 and 106 of the subject application for the reasons discussed above. It is submitted that the final rejection of claims 75, 96, and 106 of the subject application under 35 U.S.C. § 103(a) as unpatentable over the Kimpton *et al.* publication, in view of the Ledwina *et al.* publication, as assertedly motivated in view of the Jeanpierre publication, and further in view of the Goelet *et al.* '712 published international application was unwarranted and should be reversed.

VII.*a.3.iii*) Dependent Claims 76 through 82 inclusive, 85 through 87 inclusive, 91 through 95 inclusive, 97, 98, 100, 102, and 107 through 115 inclusive

Each of dependent claims 76 through 82 inclusive, 85 through 87 inclusive, 91 through 95 inclusive, 97, 98, 100, 102, and 107 through 115 inclusive respectively depends directly or indirectly on one of independent claims 75, 96 and 106 and consequently incorporates the limitations of one of independent claims 75, 96 and 106 by reference. The reasoning set forth above concerning distinctions between the Kimpton *et al.* publication considered alone or in combination with the Ledwina *et al.* publication, the Jeanpierre publication, and/or the Goelet *et al.* '712 published international application and the respective methods of independent claims 75, 96 and 106 therefore applies with equal force with respect to dependent claims 76 through 82 inclusive, 85 through 87 inclusive, 91 through 95 inclusive, 97, 98, 100, 102, and 107 through 115 inclusive. Consequently, it is submitted that the Kimpton *et al.* publication considered alone or in any combination with the Ledwina *et al.* publication, the Jeanpierre publication and/or the Goelet *et al.* '712 published international application would have neither disclosed nor in any

way suggested the subject matter of any of claims 76 through 82 inclusive, 85 through 87 inclusive, 91 through 95 inclusive, 97, 98, 100, 102, and 107 through 115 inclusive to a person of ordinary skill in the art as of the effective filing date of the subject application. It is submitted, therefore, that the final rejection under 35 U.S.C. § 103(a) of dependent claims 76 through 82 inclusive, 85 through 87 inclusive, 91 through 95 inclusive, 97, 98, 100, 102, and 107 through 115 inclusive of the subject application as unpatentable over the Kimpton *et al.* publication in view of the Ledwina *et al.* publication and further as assertedly motivated in view of the Jeanpierre publication and further in view the Goelet *et al.* '712 published international application of was without justification and should be reversed.

VII.*b*) The Rejection Under 35 U.S.C. § 112,
Second Paragraph

In the Office Action on appeal, claim 94 was finally rejected under 35 U.S.C. § 112, second paragraph, with the assertion that the expression "including such amplification by a polymerase chain reaction or a ligase chain reaction" in the expression "assaying for the given allele using genetic bit analysis, allele-specific hybridization, or allele-specific amplification, including such amplification by a polymerase chain reaction or a ligase chain reaction" recited in the claim was vague and indefinite and lacked clear antecedent basis. We submit that the antecedent basis of "including such amplification" in the immediately preceding term "allele-specific amplification," the only prior use of the word "amplification" in the claim, would have been immediately recognized by a person of ordinary skill in the art as a matter of basic English-language sentence construction and that the claim as presently worded is entirely clear and definite. Compare *MercExchange* v. *eBay*, 74 USPQ2d 1225, 1237 (Fed. Cir. 2005), vacated and remanded with respect to a separate issue, 126 S.Ct. 1837 (15 May 2006). The final rejection of claim 94 under 35 U.S.C. § 112, second paragraph, was without basis and should be reversed.

VII.*c*) Full Support for the Claims of the Subject Application Is Provided by Claimed Parent Application 08/173,173

In the seventh paragraph on page 2 of the Office Action on appeal, the claim of priority of the subject application back to application 08/173,173, filed 23 December 1993 ("the '173 application") and certain earlier applications was noted. It was asserted that it could not be determined whether the applications in question provided support for the current claims of the subject application and the claims were given the effective date of the immediate parent application 09/088,820, filed 2 June 1998 ("the '820 application").

The '820 immediate parent application is a continuation of application 08/362,266, filed 22 December 1994 ("the '266 application") and thus shares the specification of the '266 application.

The '266 application in turn is a continuation-in-part of the '173 application. Review of the specification of the '173 application as filed will show that the language of the specification, while not identical to the language of the shared specifications of the '266 application and the '820 application, generally closely parallels the language of the '266 and '820 applications. It is submitted that the pending claims of the subject application find full support in the '173 application as filed and that each claim is entitled to the benefit of the 23 December 1993 filing date of the '173 application.

It is submitted that it was error in the Office Action on appeal to treat the claims of the subject application as limited to the filing date of the '820 application and not to accord the claims the benefit of the 23 December 1993 filing date of the '173 application

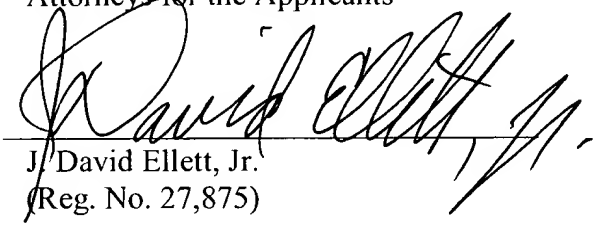
VIII. Conclusion

For the reasons set forth above, it is submitted that the claims of the subject application are patentable over the art of record considered alone or in any combination and fully meet the standards of 35 U.S.C. § 112, second paragraph, and that the claims of the application find full

support in claimed parent application 08/173,173. Reversal of the final rejections, affirmation of the support of the claims by parent application 08/173,173, and allowance of the application is therefore earnestly solicited.

Respectfully submitted,

Attorneys for the Applicants

by: 

J. David Ellett, Jr.
(Reg. No. 27,875)

Telephone No. (212) 813-1600

Claims Appendix

75. A method of determining the genotype of a subject at a locus within genetic material obtained from a biological sample from the subject, the method comprising:

A. reacting the material at the locus to produce a first reaction value indicative of the presence of a given allele at the locus;

B. forming a data set including the first reaction value;

C. establishing a distribution set of probability distributions, including at least one distribution, associating hypothetical reaction values with corresponding probabilities for each genotype of interest at the locus;

D. applying the first reaction value to each pertinent probability distribution to determine a measure of a conditional probability of each genotype of interest at the locus; and

E. determining the genotype based on the data obtained from step (D).

76. A method according to claim 75, wherein the distribution set includes a plurality of probability distributions for a corresponding plurality of genotypes of interest.

78. A method according to claim 76, further comprising:

(i) reacting the material at the locus to produce a second reaction value;

(ii) applying the first and second reaction values to each pertinent distribution to determine the probability of each genotype at the locus; and

(iii) applying the first and second reaction values to each pertinent distribution to determine a measure of the conditional probability of each genotype at the locus.

79. A method according to claim 78, wherein each probability distribution associates a hypothetical pair of first and second reaction values with a single probability of each genotype of interest.

80. A method according to claim 75, wherein:

step (B) includes the step of including in the data set other reaction values obtained under conditions comparable to those under which the first reaction value was produced; and

step (C) includes the step of using the reaction values in the data set to establish the probability distributions;

the method further comprising:

performing steps (D) and (E) with respect to each of the reaction values.

81. A method, according to claim 80, of determining the genotype at a locus within genetic material obtained from each of a plurality of samples, the method further comprising:

(i) performing step (A) with respect to the locus of material obtained from each sample;

(ii) in step (B), including in the data set reaction values obtained from each sample.

82. A method according to claim 80, of determining the genotype of selected loci within genetic material obtained from a sample, the method further comprising:

(i) performing step (A) at each of the selected loci;

(ii)in step (B), including in the data set reaction values obtained from each of the selected loci.

85. A method according to claim 75, wherein step (E) further includes the step of calculating a confidence score, associated with the determination of the genotype in step (E), based on data obtained from step (D).

86. A method according to claim 80, wherein step (E) further includes the step of calculating a confidence score, associated with the determination of the genotype in step (E), based on data from step (D), the method further comprising:

(F) determining whether a significant downward trend in confidence scores has occurred, and, in such event, entering an alarm condition.

87. A method according to claim 75, wherein each allele is defined by a single specific nucleotide.

91. A method according to claim 78, wherein step (B) includes the step of including in the data set reaction values from prior tests at the locus obtained under comparable conditions.

92. A method according to claim 82, wherein the loci are selected on the basis of their ability to discriminate among subjects.

93. A method, according to claim 77, wherein the step A' of reacting the material involves using a different reaction from that of step A and the second allele is different from the given allele.

94. A method according to claim 75, wherein step (A) includes the step of assaying for the given allele using genetic bit analysis, allele-specific hybridization, or allele-specific amplification, including such amplification by a polymerase chain reaction or a ligase chain reaction.

95. A method according to claim 82, wherein the loci are proximal to one another, so that the set of genotypes so produced may indicate a sequence of nucleotides associated with the genetic material.

96. A method of associating with a sample of genetic material from a subject one of a predetermined plurality of genotypic classes defined with respect to a genetic locus sited in the genetic material together with a corresponding confidence measure, each genotypic class

identifying either a possible genotype for the subject defined with respect to the genetic locus or a failed-experiment condition, each genotype being defined by the identity of one or more alleles defined with respect to the genetic locus, the method comprising the steps of:

(a) carrying out one or more allele-sensitive reactions on the genetic material of the sample at the genetic locus to obtain a plurality of quantitative allele-indicative reaction values, each allele-indicative reaction value being indicative of the likely presence or absence of a particular allele defined with respect to the genetic locus, the plurality of reaction values corresponding to the sample being assembled as a reaction-value data point;

(b) obtaining with respect to each of the genotypic classes corresponding reaction-value data-point conditional-probability-measure distribution information providing, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the reaction values of each hypothetical reaction-value data point given the genotypic class;

(c) evaluating for each of the genotypic classes the corresponding reaction-value data-point conditional-probability-measure distribution information with respect to the allele-indicative reaction values of the reaction-value data point corresponding to the sample to obtain for each of the corresponding genotypic classes a reaction-value data-point conditional probability measure of the reaction-value data point given the genotypic class;

(d) establishing with respect to each genotypic class a genotypic-class conditional probability measure of the genotypic class given the reaction-value data point corresponding to the sample from the reaction-value data-point conditional probability measure of the reaction-value data point given the genotypic class to obtain a set of genotypic-class conditional probability measures of the respective genotypic classes given the reaction-value data point;

(e) selecting a highest-probability genotypic-class conditional probability measure of a genotypic class given the reaction-value data point having a highest probability value from the set of genotypic-class conditional probabilities to identify a most-likely genotypic class corresponding to said highest-probability genotypic-class conditional probability measure for association with the sample; and

(f) establishing a confidence measure for the association of the most-likely genotypic class with the sample, the confidence measure being established from values of the genotypic-class conditional probability measures of the respective genotypic classes given the reaction-value data point corresponding to the sample.

97. The method according to claim 96 in which at least one genotypic class identifies a failed-experiment condition.

98. The method according to claim 96 in which the confidence measure is an LOD score.

100. The method according to claim 96 in which the allele-indicative reaction values are obtained by assaying for alleles using a method selected from the group consisting of genetic bit analysis, allele-specific hybridization, and allele-specific amplification, including such amplification by a polymerase chain reaction or a ligase chain reaction.

102. The method according to claim 98, further comprising the step of triggering an alarm condition upon detecting a statistically significant downward trend in confidence scores over time.

106. A method of associating with a sample of genetic material from a subject (i) one of a predetermined plurality of genotypic classes defined with respect to a genetic locus sited in the genetic material and (ii) a confidence measure for the association of said genotypic class with the sample, each genotypic class identifying either a possible genotype for the subject defined with respect to the genetic locus or a failed-experiment condition, each genotype being defined by the identity of one or more alleles defined with respect to the genetic locus, the method comprising the steps of:

(a) carrying out one or more allele-sensitive reactions on the genetic material of the sample at the genetic locus to obtain at least two quantitative allele-indicative reaction values, each allele-indicative reaction value being indicative of the likely presence or absence of a particular allele defined with respect to the genetic locus, the reaction values corresponding to the sample being processed to form a reaction-value data point;

(b) associating one of the genotypic classes with the sample using the reaction-value data point corresponding to the sample to define a sample genotypic class;

(c) obtaining with respect to each of the genotypic classes corresponding reaction-value data-point conditional-probability-measure distribution information providing, over a set of hypothetical reaction-value data points, a conditional probability measure as a function of the hypothetical reaction-value data point given the genotypic class;

(d) evaluating the reaction-value data-point conditional-probability-measure distribution information corresponding to the genotypic class associated with the sample with respect to the reaction-value data point corresponding to the sample to obtain a reaction-value data-point conditional probability measure of the reaction-value data point given the sample genotypic class; and

(e) establishing a confidence measure for the association of the sample genotypic class with the sample using the reaction-value data-point conditional probability measure of the reaction-value data point given the sample genotypic class.

107. The method according to claim 106 in which at least one genotypic class identifies a failed-experiment condition.

108. The method according to claim 106 in which the confidence measure is an LOD score defining a confidence score.

109. The method according to claim 108, further comprising the step of triggering an alarm condition upon detecting a statistically significant downward trend in confidence scores over time.

110. The method according to claim 106 in which the allele-indicative reaction values are obtained by assaying for alleles using a method selected from the group consisting of genetic bit analysis, allele-specific hybridization, and allele-specific amplification, including such amplification by a polymerase chain reaction or a ligase chain reaction.

111. The method according to claim 106 wherein processing the reaction values corresponding to the sample to form a reaction-value data point includes one or more of the steps of normalizing input data, subtracting background values from input data, and removing apparent outlier points.

112. The method according to claim 106 wherein the step (c) of obtaining with respect to each genotypic class corresponding reaction-value data-point conditional-probability-measure distribution information comprises, for each genotype of the genotypic classes, fitting the reaction-value data-point conditional-probability-measure distribution information corresponding to the genotype to a subset of certain input data assumed to be of the genotypic class defined with respect to such genotype.

113. A method according to claim 106, wherein the reaction values are measurements of an optical signal or a digital image intensity value.

114. A method according to claim 106, wherein the reaction values are obtained by assaying for one or more alleles in genetic material from a subject that provide information relating to a trait.

115. A method according to claim 106, wherein the reaction values are obtained by assaying for one or more alleles in genetic material from a subject that provide information pertaining to the identity or parentage of the subject.

Evidence Appendix

An Introduction to Genetic Analysis, Fifth Edition, by Griffiths *et al.* (W. H. Freeman, 1993):

A) Pages 19 through 42 and 737 through 749, entered with an information disclosure statement filed 8 April 2004; and

B) Pages 750 through 752, entered with an information disclosure statement filed 12 July 2005.

An Introduction to Genetic Analysis

Fifth Edition

Anthony J. F. Griffiths
University of British Columbia

Jeffrey H. Miller
University of California, Los Angeles

David T. Suzuki
University of British Columbia

Richard C. Lewontin
Harvard University

William M. Gelbart
Harvard University

W. H. Freeman and Company / New York

The Cover

One of the most exciting observations in modern genetics is that quite divergent animals use the same types of master regulatory genes to develop body plans that are very different in structure. The Hox (homeobox) gene cluster in vertebrates and the HOM-C (homeotic complex) cluster in insects are evolutionarily conserved sets of genes that control the individual identities of the body segments from head to tail. By classical genetic and recombinant DNA–based "gene knockout" experiments, it has been possible to demonstrate that the Hox and HOM-C genes function in parallel ways: when either a Hox or HOM-C gene is inactivated, a posteriorly located segment is converted to a more anterior identity. Normally, the fruit fly *Drosophila* has only one pair of wings, coming from the second thoracic segment. When one of the HOM-C genes is inactivated, the third thoracic segment also generates a pair of wings, producing a four-winged fly. In the mouse, the lumbar vertebrae do not have ribs. When one of the Hox genes is inactivated, the first lumbar vertebra (circled in the skeleton) is transformed into a thoracic vertebra bearing ribs. (See Chapter 22 for details.) Cover illustration by Neil Brennan, copyright 1993.

Library of Congress Cataloging-in-Publication Data

An Introduction to genetic analysis. — 5th ed. / Anthony J.F. Griffiths . . . [et al.]
p. cm.
Includes bibliographical references and index.
ISBN 0-7167-2285-2
1. Genetics. 2. Genetics—Methodology. I. Griffiths, Anthony J. F.
[DNLM: 1. Genetics. QH 430 I608]
QH430.I62 1993
575.1—dc20
DNLM/DLC
for Library of Congress 92-48977
CIP

Printed in the United States of America

2 3 4 5 6 7 8 9 0 RRD 9 9 8 7 6 5 4 3

2

Mendelian Analysis

Flowers of *Pisum sativum* (garden pea), the experimental organism used by Gregor Mendel, founder of the science of genetics. (Jeremy Burgess/Science Photo Library/Photo Researchers)

KEY CONCEPTS

The existence of genes can be inferred by observing certain progeny ratios in crosses between hereditary variants.

A discrete character difference is often determined by a difference in a single gene.

In higher organisms, each gene is represented twice in each cell.

During sex-cell formation, each member of a gene pair separates into one-half of the sex cells.

During sex-cell formation, different genes are often observed to behave independently of one another.

The gene, the basic functional unit of heredity, is the focal point of the discipline of modern genetics. In all lines of genetic research, the gene provides the common unifying thread to a great diversity of experimentation. Geneticists are concerned with the transmission of genes from generation to generation, with the physical structure of genes, with the variation in genes, and with the ways in which genes dictate the features of a species.

In this chapter we trace how the concept of the gene arose. We shall see that genetics is, in one sense, an abstract science: most of its entities began as hypothetical constructs in the minds of geneticists and were later identified in physical form.

The concept of the gene (but not the word) was first proposed in 1865 by Gregor Mendel. Until then, little progress had been made in understanding heredity. The prevailing notion was that the spermatozoon and egg contained a sampling of essences from the various parts of the parental body; at conception, these essences somehow blended to form the pattern for the new individual. This idea of **blending inheritance** evolved to account for the fact that offspring typically show some characteristics that are similar to those of both parents. However, there are some obvious problems associated with this idea, one of which is that offspring are not always an intermediate blend of their parents' characteristics. Attempts to expand and improve this theory led to no better understanding of heredity.

As a result of his research with pea plants, Mendel proposed instead a theory of **particulate inheritance.** According to Mendel's theory, characters are determined by discrete units that are inherited intact down through the generations. This model explained many observations that could not be explained by the idea of blending inheritance. It also served well as a framework for the later, more detailed understanding of the mechanism of heredity.

The importance of Mendel's ideas was not recognized until about 1900 (after his death). His written work was then rediscovered by three scientists, after each had independently obtained the same kind of results. Mendel's work constitutes the prototype for genetic analysis. He laid down an experimental and logical approach to heredity that is still used today.

Mendel's Experiments

Mendel's studies provide an outstanding example of good scientific technique. He chose research material well suited to the study of the problem at hand, designed his experiments carefully, collected large amounts of data, and used mathematical analysis to show that the results were consistent with his explanatory hypothesis. The predictions of the hypothesis were then tested in a new round of experimentation.

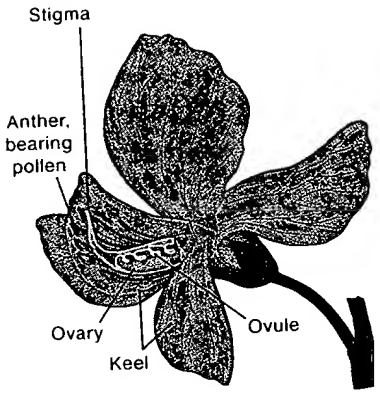


Figure 2-1 A pea flower with the keel cut and opened to expose the reproductive parts. The ovary is also shown in a cutaway view. (After J. B. Hill, H. W. Popp, and A. R. Grove, Jr., *Botany*. Copyright 1967 by McGraw-Hill.)

Mendel studied the garden pea *(Pisum sativum)* for two main reasons. First, peas were available from seed merchants in a wide array of distinct shapes and colors that could be easily identified and analyzed. Second, peas can either **self** (self-pollinate) or be cross-pollinated. The peas self because the male parts (anthers) and female parts (ovaries) of the flower — which produce the pollen containing the sperm and the ovules containing eggs, respectively — are enclosed by two petals fused to form a compartment called a keel (Figure 2-1). The gardener or experimenter can **cross** (cross-pollinate) any two pea plants at will. The anthers from one plant are removed before they have opened to shed their pollen, an operation called emasculation that is done to prevent selfing. Pollen from the other plant is then transferred to the receptive stigma with a paintbrush or on anthers themselves (Figure 2-2). Thus, the experimenter can readily choose to self or to cross the pea plants.

Other practical reasons for Mendel's choice of peas were that they are cheap and easy to obtain, take up little space, have a relatively short generation time, and produce many offspring. Such considerations enter into the choice of organism for any piece of genetic research. The choice of organism is a crucial decision and is often based on not only scientific criteria but also a good measure of expediency.

Plants Differing in One Character

Mendel chose several *characters* to study. Here, the word **character** means a specific property of an organism; geneticists use this term as a synonym for characteristic or trait.

For each of the characters he chose, Mendel obtained lines of plants, which he grew for two years to make sure they were pure. A **pure line** is a population that breeds

Figure 2-2 One technique of artificial cross-pollination, demonstrated with *Mimulus guttatus,* the yellow monkey flower. To transfer pollen, the experimenter touches anthers from the male parent to the stigma of an emasculated flower, which acts as the female parent. (Anthony Griffiths)

true for, or shows no variation in, the particular character being studied; that is, all offspring produced by selfing or crossing within the population show the same form of this character. By making sure his lines bred true, Mendel had made a clever beginning: he had established a fixed baseline for his future studies so that any changes observed following deliberate manipulation in his research would be scientifically meaningful; in effect, he had set up a control experiment.

Two of the pea lines Mendel grew proved to breed true for the character of flower color. One line bred true for purple flowers; the other, for white flowers. Any plant in the purple-flowered line — when selfed or when crossed with others from the same line — produced seeds that all grew into plants with purple flowers. When these plants in turn were selfed or crossed within the line, their progeny also had purple flowers, and so on. The white-flowered line similarly produced only white flowers through all generations. Mendel obtained seven pairs of pure lines for seven characters, with each pair differing in only one character (Figure 2-3).

Each pair of Mendel's plant lines can be said to show a **character difference** — a contrasting difference between two lines of organisms (or between two organisms) in one particular character. The differing lines (or individuals) represent different forms that the character may take: they can be called character forms, character variants, or **phenotypes.** The term phenotype (derived from Greek) literally means "the form that is shown"; it

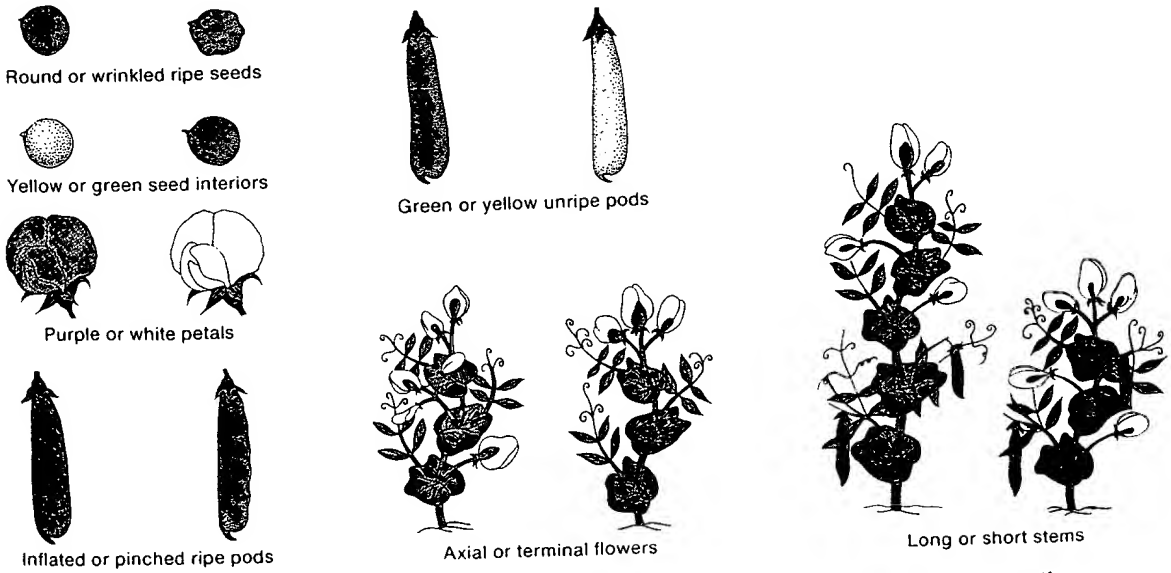


Figure 2-3 The seven character differences studied by Mendel (After S. Singer and H. Hilgard, *The Biology of People.* Copyright 1978 by W. H. Freeman and Company.)

is the term used by geneticists today. Even though such words as gene and phenotype were not coined or used by Mendel, we shall use them in describing Mendel's results and hypotheses.

Figure 2-3 shows seven pea characters, each represented by two contrasting phenotypes. Contrasting phenotypes for a particular character are the starting point for any genetic analysis. This illustrates the point made in Chapter 1 that variation is the raw material for any genetic analysis. Of course, the delineation of characters is somewhat arbitrary; an organism may be "split up" into characters in many different ways. For example, we can state one character difference of the pea plants in at least three ways:

Character	Phenotypes
flower color	purple versus white
flower purpleness	presence versus absence
flower whiteness	absence versus presence

In many cases, the description chosen is a matter of convenience (or chance). Fortunately, the choice does not alter the final conclusions of the analysis, except in the words used.

We turn now to some of Mendel's experiments with the lines breeding true for flower color. In one of his early experiments, Mendel pollinated a purple-flowered plant with pollen from a white-flowered plant. We call the plants from the pure lines the **parental generation** (P). All the plants resulting from this cross had purple flowers (Figure 2-4). This progeny generation is called the **first filial generation** (F_1). (The subsequent generations produced by selfing are symbolized F_2, F_3, and so on.)

Mendel made **reciprocal crosses**. In most plants, any cross can be made in two ways, depending on which phenotype is used as male (♂) or female (♀). For example, the two crosses

phenotype A ♀ × phenotype B ♂

phenotype B ♀ × phenotype A ♂

are reciprocal crosses. Mendel's reciprocal cross in which he pollinated a white flower with pollen from a purple-flowered plant produced the same result (all purple flowers) in the F_1 (Figure 2-5). Mendel concluded that it makes no difference which way the cross is made. If one pure-breeding parent is purple-flowered and the other is white-flowered, all plants in the F_1 have purple flowers. The purple flower color in the F_1 generation is identical to that in the purple-flowered parental plants. In this case, the inheritance obviously is not a simple blending of

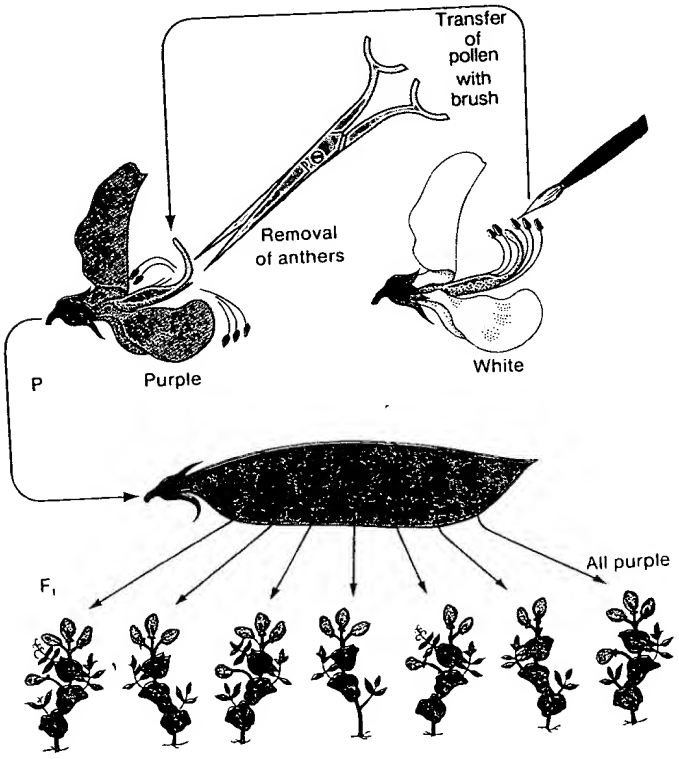


Figure 2-4 Mendel's cross of purple-flowered ♀ × white-flowered ♂.

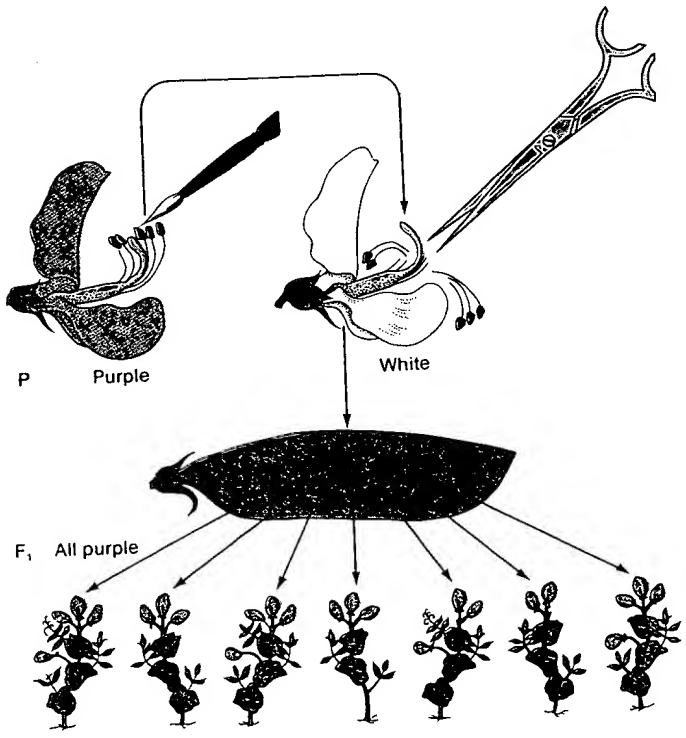


Figure 2-5 Mendel's cross of white-flowered ♀ × purple-flowered ♂.

Table 2-1. Results of All Mendel's Crosses in Which Parents Differed for One Character

Parental phenotypes	F_1	F_2	F_2 Ratio
1. Round × wrinkled seeds	All round	5474 round; 1850 wrinkled	2.96:1
2. Yellow × green seeds	All yellow	6022 yellow; 2001 green	3.01:1
3. Purple × white petals	All purple	705 purple; 224 white	3.15:1
4. Inflated × pinched pods	All inflated	882 inflated; 299 pinched	2.95:1
5. Green × yellow pods	All green	428 green; 152 yellow	2.82:1
6. Axial × terminal flowers	All axial	651 axial; 207 terminal	3.14:1
7. Long × short stems	All long	787 long; 277 short	2.84:1

purple and white colors to produce some intermediate color. To maintain a theory of blending inheritance, we would have to assume that the purple color is somehow "stronger" than the white color and completely overwhelms any trace of the white phenotype in the blend.

Next, Mendel selfed the F_1 plants, allowing the pollen of each flower to fall on its own stigma. He obtained 929 pea seeds from this selfing (the F_2 individuals) and planted them. Interestingly, some of the resulting plants were white flowered; the white phenotype had reappeared. Mendel then did something that, more than anything else, marks the birth of modern genetics: he *counted* the numbers of plants with each phenotype. This procedure had seldom, if ever, been used in genetic studies before Mendel's work. Indeed, others had obtained remarkably similar results in breeding studies but had failed to count the numbers in each class. Mendel counted 705 purple-flowered plants and 224 white-flowered plants. He noted that the ratio of 705:224 is almost a 3:1 ratio (in fact, it is 3.1:1).

Mendel repeated the crossing procedures for the six other pairs of pea character differences. He found the same 3:1 ratio in the F_2 generation for each pair (Table 2-1). By this time, he was undoubtedly beginning to believe in the significance of the 3:1 ratio and to seek an explanation for it. In all cases, one parental phenotype disappeared in the F_1 and reappeared in one-fourth of the F_2. The white phenotype, for example, was completely absent from the F_1 generation but reappeared (in its full original form) in one-fourth of the F_2 plants.

It is very difficult to devise an explanation of this result in terms of blending inheritance. Even though the F_1 flowers were purple, the plants evidently still carried the *potential* to produce progeny with white flowers. Mendel inferred that the F_1 plants receive from their parents the ability to produce both the purple phenotype and the white phenotype and that these abilities are retained and passed on to future generations rather than blended. Why is the white phenotype not expressed in the F_1 plants? Mendel used the terms **dominant** and **recessive** to describe this phenomenon without explaining the mechanism. In modern terms, the purple phenotype is dominant to the white phenotype and the white phenotype is recessive to purple. Thus the operational definition of dominance is provided by the phenotype of an F_1 established by intercrossing two pure lines. The parental phenotype that is expressed in such F_1 individuals is by definition the dominant phenotype.

Mendel went on to show that in the class of F_2 individuals showing the dominant phenotype there were in fact two genetically distinct subclasses. In this case he was working with seed color. In peas the color of the seed is determined by the genetic constitution of the seed itself, and not by the maternal parent as in some plant species. This is convenient because the investigator can treat each pea as an individual and can observe its phenotype directly without having to grow up a plant from it as must be done for flower color. This also means much larger numbers can be examined, and it is more convenient to extend studies into subsequent generations. The seed colors Mendel used were yellow and green. He crossed a pure yellow line with a pure green line, and observed that the F_1 peas that appeared were all yellow. Symbolically

$$\begin{array}{lc} P & \text{yellow} \times \text{green} \\ & \downarrow \\ F_1 & \text{all yellow} \end{array}$$

Therefore, by definition, yellow is the dominant phenotype and green is recessive.

Mendel grew F_1 plants from these F_1 peas, and then selfed the plants. The peas that developed on the F_1 plants represented the F_2 generation. He observed that in the pods of the F_1 plants $\frac{3}{4}$ of the peas were yellow and $\frac{1}{4}$ were green. Here again in the F_2 we see a 3:1 phenotypic ratio. Mendel took a sample consisting of 519 yellow F_2 peas and grew plants from them. These F_2 plants were selfed individually and the peas that developed were noted. Mendel found that 166 of the plants bore only yellow peas, and each of the remaining 353 plants bore a mixture of yellow and green peas in a 3:1 ratio. In addition, green F_2 peas were grown up and selfed, and were found to bear only green peas. In summary, all of the F_2 greens were

evidently pure breeding like the green parental line, but of the F_2 yellows $\frac{2}{3}$ were like the F_1 yellows (producing yellow and green seeds in a 3:1 ratio) and $\frac{1}{3}$ were like the pure-breeding yellow parent. Thus, the study of the next generation (the F_3) revealed that underlying the 3:1 phenotypic ratio in the F_2 generation there was a more fundamental 1:2:1 ratio:

$\frac{3}{4}$ yellow → $\frac{1}{4}$ pure-breeding yellow
$\frac{3}{4}$ yellow → $\frac{2}{4}$ "impure" yellow
$\frac{1}{4}$ green → $\frac{1}{4}$ pure-breeding green

Further studies showed that such 1:2:1 ratios underlay all of the phenotypic ratios that Mendel had observed. Thus, the problem really was to explain the 1:2:1 ratio. Mendel's explanation was a classic example of a creative model or hypothesis derived from observation and well suited for testing by further experimentation. He deduced the following explanation of the 1:2:1 ratio.

1. There are hereditary determinants of a particulate nature. (Mendel saw no blending of phenotypes, so he was forced to draw this conclusion.) We now call these determinants genes.
2. Each adult pea plant has two genes—a **gene pair**—in each cell for each character studied. Mendel's reasoning here was obvious: the F_1 plants, for example, must have had one gene that was responsible for the dominant phenotype and another gene that was responsible for the recessive phenotype, which showed up only in later generations.
3. The members of the gene pairs segregate (separate) equally into the gametes, or eggs and sperm.
4. Consequently, each gamete carries only one member of each gene pair.
5. The union of one gamete from each parent to form the first cell (or zygote) of a new progeny individual is random—that is, gametes combine without regard to which member of a gene pair is carried.

These points can be illustrated diagrammatically for a general case, using *A* to represent the gene that determines the dominant phenotype and *a* to represent the gene for the recessive (as Mendel did). This is similar to the way a mathematician uses symbols to represent abstract entities of various kinds. In Figure 2-6, these symbols are used to illustrate how the above five points explain the 1:2:1 ratio.

The whole model made logical sense of the data. However, many beautiful models have been knocked down under test. Mendel's next job was to test his model. He did this by taking (for example) an F_1 plant that grew from a yellow seed and crossing it with a plant grown from a green seed. A 1:1 ratio of yellow to green seeds could be predicted in the next generation. If we let *Y* stand for the gene that determines the dominant phenotype (yellow seeds) and *y* stand for the gene that determines the recessive phenotype (green seeds), we can diagram Mendel's predictions, as shown in Figure 2-7. In this experiment, he obtained 58 yellow (*Yy*) and 52 green (*yy*), a very close approximation to the predicted 1:1 ratio and confirmation of the equal segregation of *Y* and *y* in the F_1 individual. This concept of **equal segregation** has been given formal recognition as Mendel's first law.

Mendel's First Law The two members of a gene pair segregate from each other into the gametes, so that one-half of the gametes carry one member of the pair and the other one-half of the gametes carry the other member of the pair.

Now we need to introduce some more terms. The individuals represented by *Aa* are called **heterozygotes,** or sometimes **hybrids,** whereas the individuals in pure lines are called **homozygotes.** In words such as these, "hetero-" means different and "homo-" means identical. Thus, an *AA* plant is said to be **homozygous dominant;** an *aa* plant is homozygous for the recessive gene, or **homozygous recessive.** As we saw in Chapter 1, the designated genetic constitution of the character or characters under study is called the **genotype.** Thus, *YY* and *Yy*, for example, are different genotypes even though the seeds of both types are of the same phenotype (that is, yellow). In such a situation the phenotype can be thought of simply as the outward manifestation of the underlying genotype. Note that underlying the 3:1 phenotypic ratio in the F_2 there is a 1:2:1 genotypic ratio of *YY*:*Yy*:*yy*.

Note that strictly speaking the expressions "dominant" and "recessive" are properties of the phenotype. The dominant phenotype is established in analysis by the

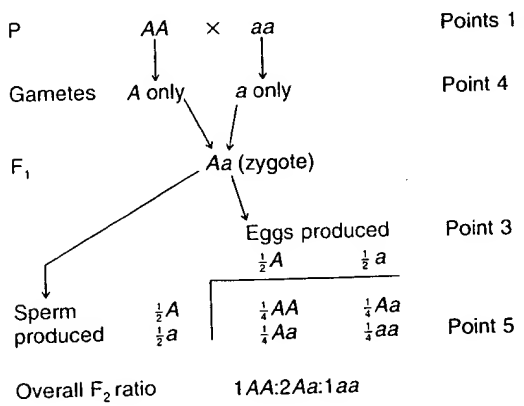


Figure 2-6 Mendel's model of the hereditary determinants of a character difference in the P, F_1, and F_2. The five points are those listed in the text.

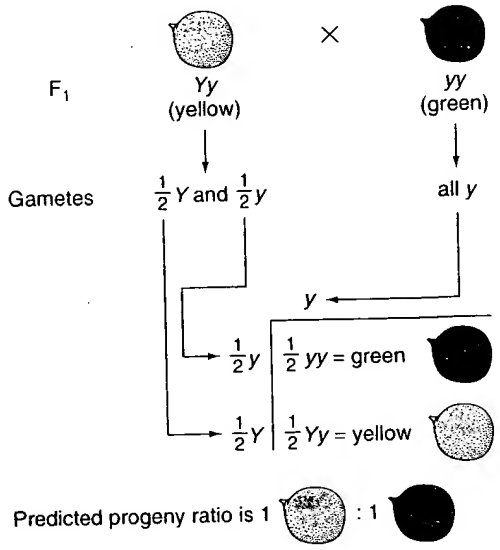


Figure 2-7 Prediction of a 1:1 phenotypic ratio in a cross of an F_1 yellow with a green in peas.

appearance of the F_1. Obviously, however, a phenotype (which is merely a description) cannot really exert dominance. Mendel showed that the dominance of one phenotype over another is in fact due to the dominance of one member of a gene pair over the other.

Let's pause to let the significance of this work sink in. What Mendel did was to develop an analytic scheme for the identification of genes regulating any biological character or function. Let's take petal color as an example. Starting with two different phenotypes (purple and white) of one character (petal color), Mendel was able to show that the difference was caused by one gene pair. Modern geneticists would say that Mendel's analysis had identified a gene for petal color. What does this mean? It means that in these organisms there is a gene that has a profound effect on the color of the petals. This gene can exist in different forms: the dominant form of the gene (represented by *C*) causes purple petals, and the recessive form of the gene (represented by *c*) causes white petals. The forms *C* and *c* are called **alleles** (or alternative forms) of that gene for petal color. They are given the same letter symbol to show that they are forms of one gene. We could express this another way by saying that there is a kind of gene, called phonetically a "see" gene, with alleles *C* and *c*. Any individual pea plant will always have two "see" genes, forming a gene pair, and the actual members of the gene pair can be *CC*, *Cc*, or *cc*. Notice that although the members of a gene pair can produce different effects, they obviously both affect the same character.

The related terms gene and allele are potentially confusing, but to clarify the concepts let us jump ahead to think about what they signify at the level of DNA. When alleles like *A* and *a* are examined at the DNA level using modern technology, it is generally found that they are identical for most of their sequence, and differ only at one or a few nucleotides out of the thousands that make up the gene. Therefore, we see that the alleles are truly different versions of the same basic gene. Looked at another way, gene is the generic term and allele is specific.

We have seen that although the terms dominant and recessive are defined at the level of phenotype, the phenotypes clearly reflect the different actions of various alleles. Therefore we can legitimately use the phrases dominant allele and recessive allele as the determinants of dominant and recessive phenotypes. (Note that this usage is more correct than the terms dominant gene and recessive gene, although these terms are sometimes used.)

How does the allele terminology relate to the gene pair concept? A gene pair can consist of identical alleles (in homozygotes) or different alleles (in heterozygotes). When a gene pair segregates as described by Mendel's first law, then it can be identical alleles that segregate, as in homozygotes, or different alleles that segregate, as in heterozygotes. Of course it will only be the heterozygous segregation that can lead to a phenotypic segregation (that is, to two separate phenotypes) in the progeny.

The basic route of Mendelian analysis for a single character is summarized in Table 2-2.

Message The existence of genes was originally inferred (and is still inferred today) by observing precise mathematical ratios in the descendants of two genetically different parental individuals.

Plants Differing in Two Characters

In the experiments described so far we have been concerned with what is sometimes called a **monohybrid cross,** the crossing or selfing of heterozygotes produced by a mating between individuals from two pure lines that differ in a single gene that controls a character difference. Now we can ask what happens in a **dihybrid cross,** in which the pure parental lines differ in two genes that control two separate character differences. We can use the same symbolism that Mendel used to indicate the genotypes of seed color (*Y* and *y*) and seed shape (*R* and *r*). *R* gives round seeds, and *r* gives wrinkled. In a monohybrid cross, a ratio of ¾ round and ¼ wrinkled is observed in the F_2 (Figure 2-8).

A pure line of *RR yy* plants, on selfing, produces seeds that are round and green. Another pure line is *rr YY*; on selfing, this line produces wrinkled yellow seeds. When Mendel crossed plants from these two lines, he obtained round, yellow F_1 seeds, as expected. The results in the F_2 are summarized in Figure 2-9. Mendel per-

Table 2-2. Summary of the Modus Operandi for Establishing Simple Mendelian Inheritance

Experimental procedure:	1. Choose pure lines showing a character difference (purple versus white flowers). 2. Cross the lines. 3. Self the F_1 individuals.
Results:	F_1 is all purple; F_2 is $\frac{3}{4}$ purple and $\frac{1}{4}$ white.
Inferences:	1. The character difference is controlled by a major gene for flower color. 2. The dominant allele of this gene causes purple petals; the recessive allele causes white petals.

Symbolic interpretation:

Character	Phenotype	Genotype	Allele	Gene
Flower color	Purple (dominant)	*CC* (homozygous dominant) *Cc* (heterozygous)	*C* (dominant) *c* (recessive)	Flower-color gene
	White (recessive)	*cc* (homozygous recessive)		

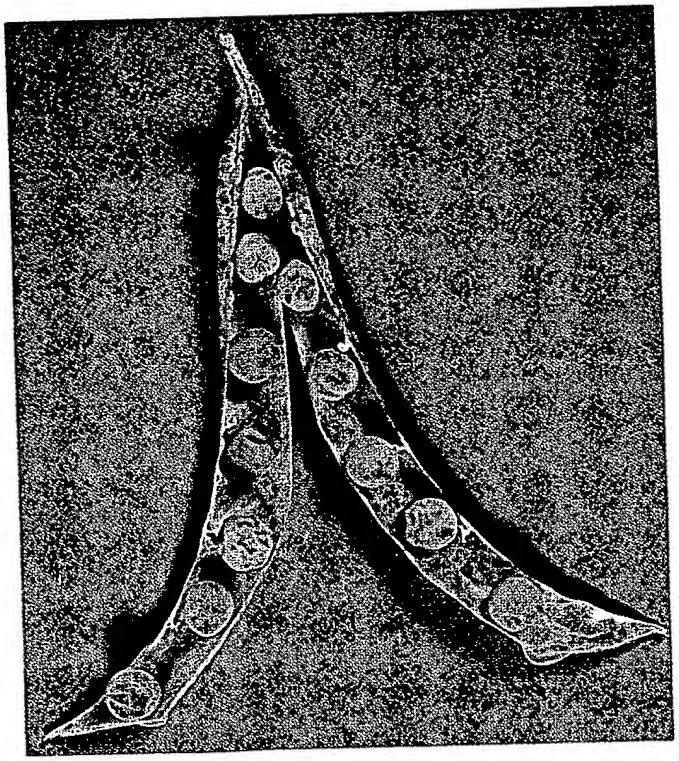


Figure 2-8 Round (*R* –) and wrinkled (*rr*) peas in a pod of a selfed heterozygous plant (*Rr*). The phenotypic ratio in this pod happens to be precisely the 3:1 ratio expected on average in the progeny of this selfing. (Recent molecular studies have shown that the wrinkled allele used by Mendel is produced by insertion into the gene of a segment of mobile DNA of the type to be discussed in Chapter 20.) (Madan K. Bhattacharyya)

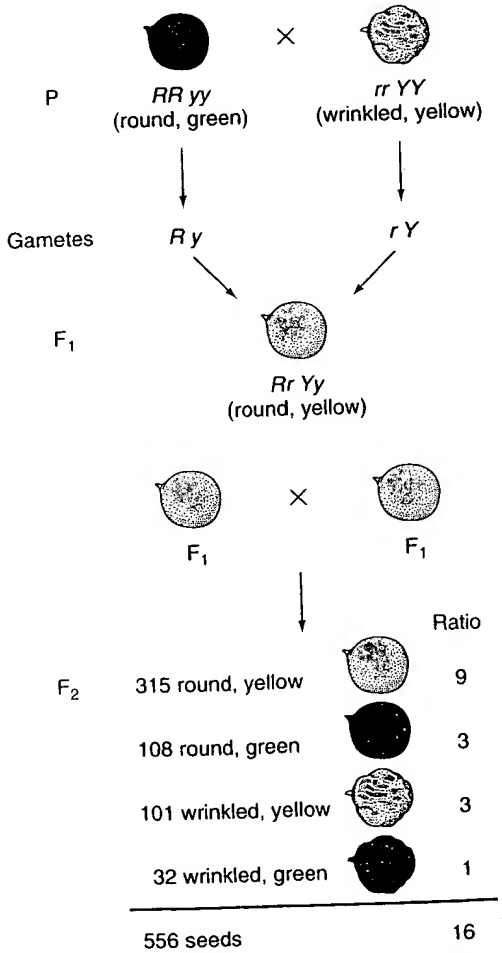


Figure 2-9 The F_2 generation resulting from a dihybrid cross.

formed similar experiments using other pairs of characters in many other dihybrid crosses; in each case, he obtained 9:3:3:1 ratios. So, he had another phenomenon to explain, and some more numbers to turn into an idea.

Mendel first checked to see whether the ratio for each gene in the dihybrid cross was the same as that for a monohybrid cross. If you look at only the round and wrinkled phenotypes and add up all the seeds falling into these two classes in Figure 2-9, the totals are $315 + 108 = 423$ round and $101 + 32 = 133$ wrinkled. Hence, the monohybrid 3:1 ratio still prevails. Similarly, the ratio of yellow seeds to green seeds is $(315 + 101):(108 + 32) = 416:140$, again very close to 3:1. From this clue, Mendel concluded that the two hereditary systems are independent. He was mathematically astute enough to realize that the 9:3:3:1 ratio is nothing more than a random combination of two independent 3:1 ratios.

This is a convenient point at which to introduce some elementary rules of probability that we will use often throughout this book.

1. *Definition of probability.*

$$\textbf{Probability} = \frac{\text{The number of times an event is expected to happen}}{\text{The number of opportunities for an event to happen (or the number of trials)}}$$

For example, the probability of rolling a 4 on a die in a single trial is written

$$p(\text{of a 4}) = \tfrac{1}{6}$$

because the die has six sides. If each side is equally likely to turn up, then on the average one 4 should turn up for each six rolls.

2. *The product rule.* The probability that two independent events will occur simultaneously is the product of their respective probabilities. For example, rolling a die twice is two independent events, and

$$p(\text{of two 4s}) = \tfrac{1}{6} \times \tfrac{1}{6} = \tfrac{1}{36}$$

3. *The sum rule.* The probability of *either one* of two independent (mutually exclusive) events is the sum of their individual probabilities. For example, when two dice are rolled together,

$$p(\text{of two 4s } \textit{or} \text{ two 5s}) = \tfrac{1}{36} + \tfrac{1}{36} = \tfrac{1}{18}$$

The composition of the F_2 from the pea dihybrid cross can be predicted if the mechanism for putting R or r into a gamete is *independent* of the mechanism for putting Y or y into the gamete. The frequency of gamete types can be calculated by determining their probabilities according to the rules just given. Thus, if you pick a gamete at random, the *probability* of picking a certain type of gamete is the same as the frequency of that type of gamete.

We know from Mendel's first law that a heterozygote produces gametes in the proportions

$$Y \text{ gametes} = y \text{ gametes} = \tfrac{1}{2}$$
$$R \text{ gametes} = r \text{ gametes} = \tfrac{1}{2}$$

An $Rr\,Yy$ plant forms four types of gametes. The probability that a gamete carries R and Y is written $p(R\,Y)$. Similarly, $p(R\,y)$ denotes the probability that a gamete carries R and y. Under the assumption that the segregation of R or r into a gamete is independent of the segregation of Y or y into that same gamete, we can use the product rule to calculate the probability of each gametic combination:

$$p(R\,Y) = \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}$$
$$p(R\,y) = \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}$$
$$p(r\,y) = \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}$$
$$p(r\,Y) = \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}$$

Thus, we can represent the F_2 generation by a grid named (after its inventor) a Punnett square, as shown in Figure 2-10. The columns of the square tabulate the contributions of the male parents to the F_2 and the rows tabulate those of the female parents.

The probability of $\frac{1}{16}$ shown for each box in the square is also obtained using the product rule. The constitution of a zygote is the result of two independent events — the event that formed the male gamete and the event that formed the female. Thus, for example, the probability (or frequency) of $RR\,YY$ zygotes (combining an $R\,Y$ male gamete with an $R\,Y$ female gamete) is $\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$. Grouping all the types that look the same from Figure 2-9, we find the 9:3:3:1 ratio (now not so mysterious) in all its beauty:

round, yellow	$\frac{9}{16}$ or 9
round, green	$\frac{3}{16}$ or 3
wrinkled, yellow	$\frac{3}{16}$ or 3
wrinkled, green	$\frac{1}{16}$ or 1

The independence of the allelic segregations for two different genes is an important concept. It is called **independent assortment,** and its general statement is now known as Mendel's second law.

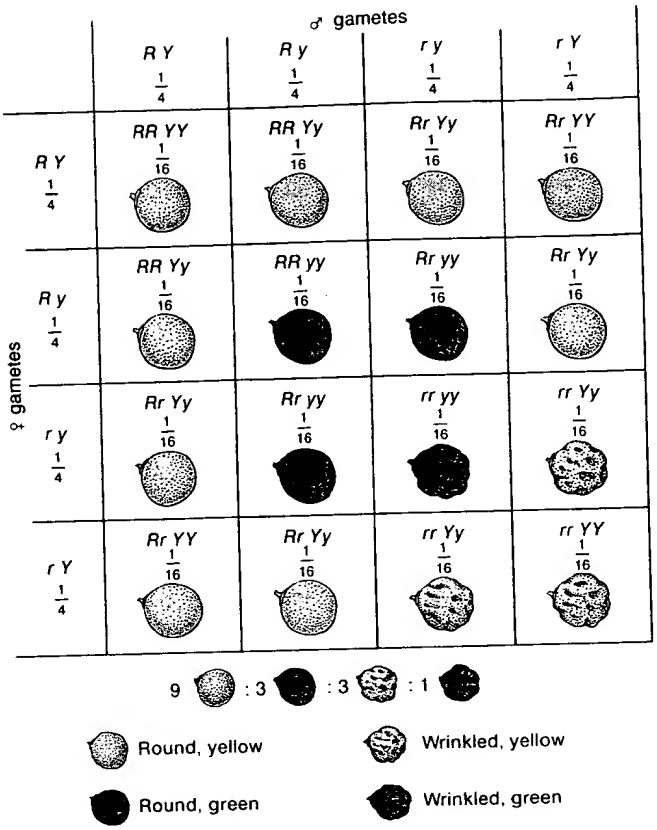


Figure 2-10 Punnett square showing predicted genetic and phenotypic constitution of the F_2 generation from the dihybrid cross shown in Figure 2-9.

> **Mendel's Second Law** During gamete formation the segregation of alleles of one gene is independent of the segregation of alleles of another gene.

A note of warning: we shall see later that a phenomenon called gene linkage results in an important exception to Mendel's second law.

Note how Mendel's counting led to the discovery of such unexpected regularities as the 9:3:3:1 ratio and how a few simple assumptions (such as equal segregation and independent assortment) can explain this ratio that initially seems so baffling. Although it was unappreciated at the time, Mendel's quantitative approach provided the key to an understanding of genetic mechanisms.

Of course, Mendel went on to test this second law too. For example, he crossed an F_1 dihybrid $Rr\,Yy$ with a doubly homozygous recessive strain $rr\,yy$. A cross to a homozygous recessive is now known as a **testcross.** Testcrosses allow the experimenter to focus on the genotype underlying a dominant phenotype in one individual because the homozygous recessive parent contributes only recessive alleles to the progeny. We shall see this kind of cross many times in this book. For his testcross, Mendel predicted that the dihybrid $Rr\,Yy$ would produce the gametic types $R\,Y$, $R\,y$, $r\,y$, and $r\,Y$ in equal frequency — that is, as shown along one edge of the Punnett square in Figure 2-10, in the frequencies $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, and $\frac{1}{4}$. On the other hand, because it is homozygous, the $rr\,yy$ plant produces only one gamete type ($r\,y$), regardless of equal segregation or independent assortment. Thus, the progeny phenotypes should reflect directly the gametic types from the $Rr\,Yy$ parent (because the recessive $r\,y$ contribution from the $rr\,yy$ parent does not alter the phenotype indicated by the other gamete). Hence, Mendel predicted a 1:1:1:1 ratio of $Rr\,Yy$, $Rr\,yy$, $rr\,yy$, and $rr\,Yy$ progeny from this testcross, and his prediction was confirmed. He tested the concept of independent assortment intensively on four different combinations of his characters and found that it applied to every combination.

Of course, the deduction of equal segregation and independent assortment as abstract concepts that explain the observed facts leads immediately to the question of what structures or forces can account for such behavior of genes. The idea of equal segregation seems to indicate that both alleles of a gene actually exist in some kind of orderly, paired configuration from which they can separate cleanly during gamete formation (Figure 2-11). If alleles of another gene behave independently in the same way, then we have independent assortment (Figure 2-12). But this is all speculation at the present stage of our discussion, as it was after the rediscovery of Mendel's work. The actual mechanisms are now known, and in Chapter 3 we shall see that it is the chromosomal location of genes that is responsible for their equal segregation and independent assortment.

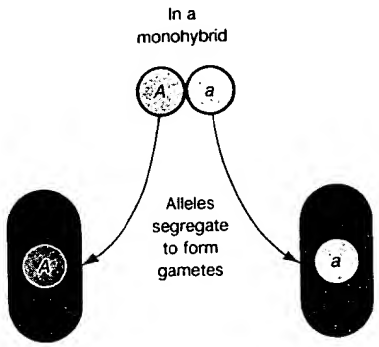


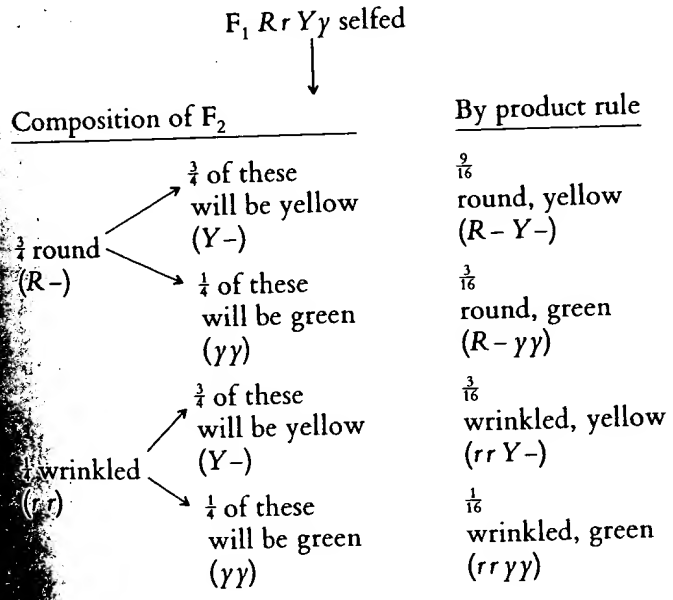
Figure 2-11 The segregation of alleles into gametes.

Message Mendel established the ground rules for genetic analysis. His work made it possible to infer the existence and nature of hereditary units and mechanisms without ever seeing them. The analysis of phenotypic frequencies in the progenies of controlled crosses is still an integral part of the experimental approach used in much of modern genetics.

Methods for Calculating Genetic Ratios

We pause here for a few words on calculating phenotypic and genotypic ratios. The Punnett square is graphic and reliable, but it is unwieldy; it is suited only for illustration, not for efficient calculation.

A branch diagram is useful for solving some problems. For example, the 9:3:3:1 phenotypic ratio can be derived by drawing a branch diagram and applying the product rule to determine frequencies. (Note the use of the convention that $R-$ represents both RR and Rr; that is, either allele can occupy the space indicated by the dash.)



The branch diagram is, of course, a graphic expression of the product rule. It can be used for phenotypic or genotypic ratios.

The diagram can be extended to a trihybrid ratio (such as $Aa\,Bb\,Cc \times Aa\,Bb\,Cc$) by drawing another set of branches on the end. However, as the number of genes increases, the number of identifiable phenotypes rises startlingly and the number of genotypes climbs even more steeply, as shown in Table 2-3. With such large numbers, even the branch method becomes unwieldy.

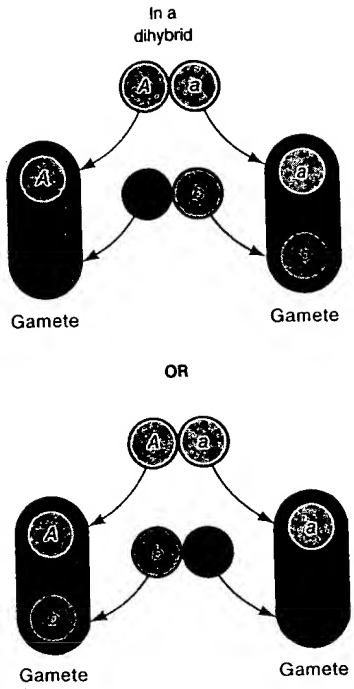


Figure 2-12 The segregation of alleles of two independent genes into gametes.

Table 2-3. Rise in Number of Phenotypes and Genotypes as the Power of the Number of Segregating Gene Pairs

Number of segregating gene Pairs	Number of phenotypes	Number of genotypes
1	2	3
2	4	9
3	8	27
4	16	81
.	.	.
.	.	.
.	.	.
n	2^n	3^n

In such cases, we must resort to devices based directly on the product and sum rules. For example, what proportion of progeny from the cross

$$Aa\,Bb\,Cc\,Dd\,Ee\,Ff \times Aa\,Bb\,Cc\,Dd\,Ee\,Ff$$

will be $AA\,bb\,Cc\,DD\,ee\,Ff$? The answer is easily obtained if the pairs of alleles all assort independently, thereby allowing use of the product rule. Thus, $\frac{1}{4}$ of the progeny will be AA, $\frac{1}{4}$ will be bb, $\frac{1}{2}$ will be Cc, $\frac{1}{4}$ will be

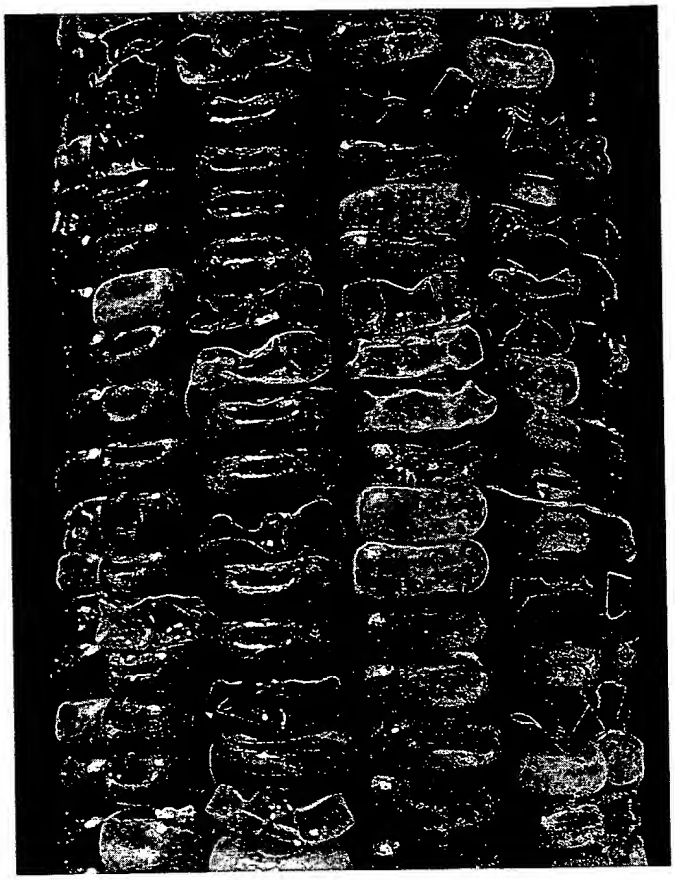


Figure 2-13 A 9:3:3:1 ratio in the phenotypes of kernels of corn. Each kernel represents a progeny individual. The progeny result from a self on an individual of genotype *Aa Bb*, where *A* = purple, *a* = yellow, *B* = smooth, and *b* = wrinkled. (Anthony Griffiths)

DD, $\frac{1}{4}$ will be *ee*, and $\frac{1}{2}$ will be *Ff*, so we obtain the answer by multiplying these frequencies:

$$p(AA\,bb\,Cc\,DD\,ee\,Ff) = \tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{2} \times \tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{2}$$
$$= \tfrac{1}{1024}$$

Let us return now to Mendel's work. When Mendel's results were rediscovered in 1900, his principles were tested in a wide spectrum of eukaryotic organisms (organisms with cells that contain nuclei). The results of these tests showed that Mendelian principles were generally applicable. Mendelian ratios (such as 3:1, 1:1, 9:3:3:1, and 1:1:1:1) were extensively reported (Figure 2-13), suggesting that equal segregation and independent assortment are fundamental hereditary processes found throughout nature. Mendel's laws are not merely laws about peas but laws about the genetics of eukaryotic organisms in general. The experimental approach used by Mendel can be extensively applied in plants. However, in some plants and in most animals, the technique of selfing is impossible. This problem can be circumvented by crossing identical genotypes. For example, an F_1 animal resulting from the mating of parents from differing pure lines can be mated to its F_1 siblings (brothers or sisters) to produce an F_2. The F_1 individuals are identical for the genes in question, so the F_1 cross is equivalent to a selfing.

Simple Mendelian Genetics in Humans

Other systems present some special problems in the application of Mendelian methodology. One of the most difficult, yet most interesting, is the human species. Obviously, controlled crosses cannot be made, so human geneticists must resort to scrutinizing records in the hope that informative matings have been made by chance. Such a scrutiny of records of matings is called **pedigree analysis.** A member of a family who first comes to the attention of a geneticist is called the **propositus.** Usually the phenotype of the propositus is exceptional in some way (for example, the propositus might be a dwarf). The investigator then traces the history of the character in the propositus back through the history of the family and draws up a family tree or pedigree, using certain standard symbols given in Figure 2-14. (The terms autosomal and sex-linked in the figure will be explained later; they are included to make the list of symbols complete.)

Many pairs of contrasting human phenotypes are determined by pairs of alleles inherited in exactly the same manner shown by Mendel's peas. Pedigree analysis can reveal such inheritance patterns, but the clues in the pedigree have to be interpreted differently depending on whether one of the contrasting phenotypes is a rare disorder or whether both phenotypes of a pair are part of normal variation. These are considered separately in the following sections.

Medical Genetics

Medicine is concerned with the disorders of human beings, and many of these disorders are inherited as dominant or recessive phenotypes in a simple Mendelian manner. Generally, the propositus, or index case, is an individual who comes to the attention of a physician as a patient. If there is a family history of the disorder, a pedigree is constructed as far as available information allows, and then the pedigree is analyzed. Let us consider the inheritance of recessives first.

The unusual condition of a recessive disorder is determined by a recessive allele, and the corresponding

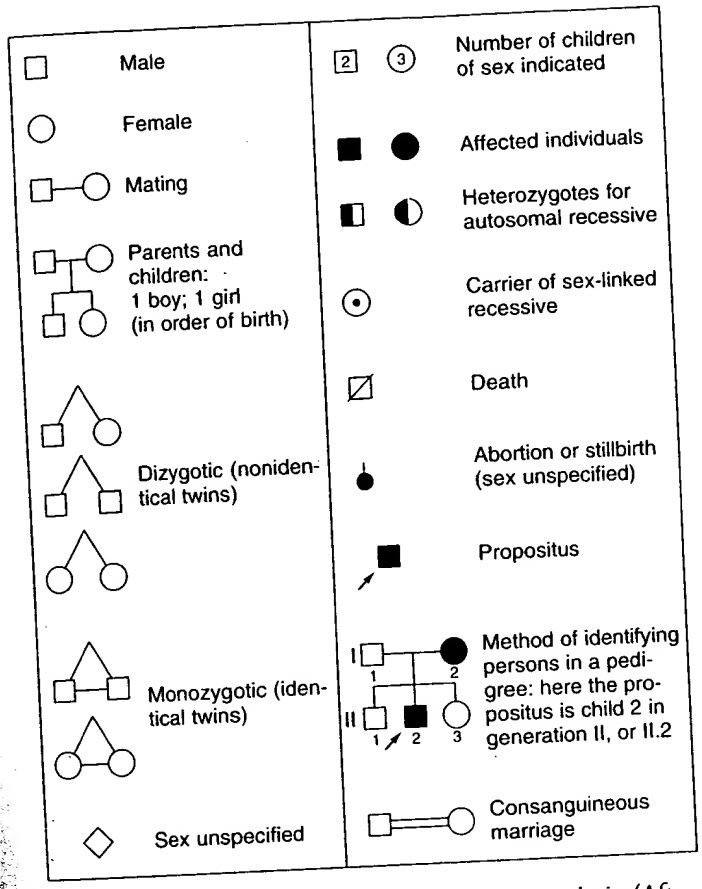
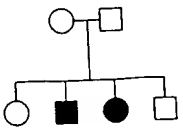
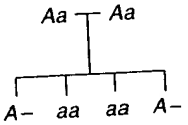


Figure 2-14 Symbols used in human pedigree analysis. (After W. F. Bodmer and L. L. Cavalli-Sforza, *Genetics, Evolution, and Man*. Copyright 1976 by W. H. Freeman and Company.)

unaffected phenotype is determined by a dominant allele. For example, the human disease phenylketonuria (PKU) is inherited in a simple Mendelian manner as a recessive phenotype, with PKU determined by an allele p and the normal condition by P. Therefore, sufferers from this disease are of genotype pp, and people who do not have the disease are either PP or Pp. What patterns in a pedigree would reveal such an inheritance? The two key points are that generally the disease appears in the progeny of unaffected parents and that the affected progeny includes both males and females. When we know that both male and female progeny are affected, we can assume that we are dealing with simple Mendelian inheritance, and not a special type of sex-linked inheritance that we will discuss in Chapter 3. The following typical pedigree illustrates the key point that affected children are born to unaffected parents:



From this pattern we can immediately deduce simple Mendelian inheritance of the recessive allele responsible for the exceptional phenotype (indicated by shading). Furthermore, we can deduce that the parents are both heterozygotes, say Aa: both must have an a allele because each contributed an a allele to each affected child, and both must have an A allele because they are phenotypically normal. We can identify the genotypes of the children (in the order shown) as $A-$, aa, aa, and $A-$. Hence, the pedigree can be rewritten



Notice another interesting feature of pedigree analysis: even though Mendelian rules are at work, Mendelian ratios are rarely observed in families because the sample size is too small. In the above example, we see a 1:1 phenotypic ratio in the progeny of a monohybrid cross. If the couple were to have, say, 20 children, the ratio would be something like 15 unaffected children and 5 with PKU (a 3:1 ratio), but in a sample of four any ratio is possible, and all ratios are commonly found.

The pedigrees of recessive disorders tend to look rather bare, with few shaded symbols. A recessive condition shows up in groups of affected siblings, and the people in earlier and later generations tend not to be affected. To understand why this is so, it is important to have some feel for the genetic structure of populations underlying such rare conditions. By definition, if the condition is rare, most people do not carry the abnormal allele. Furthermore, most of those people who do have the abnormal allele are heterozygous for it rather than homozygous. The basic reason that heterozygotes are much more common than recessive homozygotes is that to be a recessive homozygote, both of your parents must have had the a gene, but to be a heterozygote all you need is one parent with the gene.

Geneticists have a quantitative way to connect the rareness of an allele with the commonness or rarity of heterozygotes and homozygotes in a population. They obtain the relative frequencies of genotypes in a population by assuming that the population is in Hardy-Weinberg equilibrium, to be fully discussed in Chapter 25. Under this simplifying assumption, if the relative proportions of two alleles A and a in a population are p and q respectively, then the frequencies of the three possible genotypes are given by p^2 for AA, $2pq$ for Aa, and q^2 for aa. A numerical example illustrates this. If we assume that the frequency, q, of a recessive, disease-causing allele is $\frac{1}{50}$, then p, is $\frac{49}{50}$, and the frequency of homozygotes with the disease is $q^2 = (\frac{1}{50})^2 = \frac{1}{2500}$, and the frequency of het-

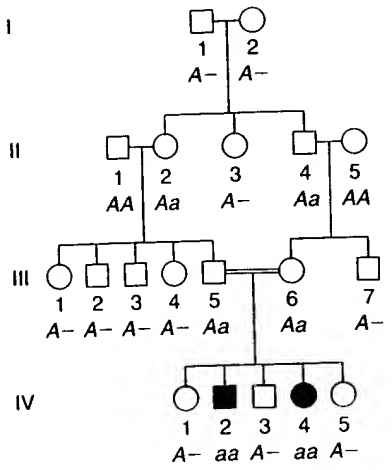


Figure 2-15 Pedigree of a rare recessive phenotype determined by a recessive allele *a*. Gene symbols normally are not included in pedigree charts, but genotypes are inserted here for reference. Note that individuals II-1 and II-5 marry into the family; they are assumed to be normal because the heritable condition under scrutiny is rare. Note also that it is not possible to be certain of the genotype in some individuals with normal phenotype; such individuals are indicated by *A*–.

erozygotes is $2pq = 2 \times \frac{49}{50} \times \frac{1}{50} =$ approximately $\frac{1}{25}$. Hence, for this example, we see that heterozygotes are 100 times more frequent than disease sufferers, and this ratio increases the rarer the allele.

The formation of an affected individual usually depends on the chance union of unrelated heterozygotes. However, inbreeding (mating between relatives) increases the chance that a mating will be between two heterozygotes. An example of a cousin marriage is shown in Figure 2-15. You can see from the figure that an ancestor who is a heterozygote may produce many descendants who are also heterozygotes. Matings between relatives thus run a higher risk of producing abnormal phenotypes caused by homozygosity for recessive alleles than do matings between non-relatives. It is for this reason that first cousin marriages contribute a large proportion of the sufferers of recessive diseases in the population.

What are some examples of recessive disorders in humans? We have already used PKU as an example of pedigree analysis, but what kind of phenotype is it? PKU is a disease in which the body cannot properly process the amino acid phenylalanine, a component of all proteins in the food we eat. This substance builds up in the body and is converted to phenylpyruvic acid, which interferes with the development of the nervous system, leading to mental retardation. Babies are now routinely tested for this processing deficiency upon birth. If the deficiency is detected, phenylalanine can be withheld by use of a special diet, and the development of the disease can be arrested.

Cystic fibrosis is another disease inherited according to Mendelian rules as a recessive phenotype. This disease has received much recent attention because the allele that causes it was isolated in 1989 and the sequence of its DNA determined. This has led to an understanding of gene function in affected and unaffected individuals, giving hope for more effective treatment. Cystic fibrosis is a disease whose most important symptom is the secretion of large amounts of mucus into the lungs, resulting in death from a combination of effects, but usually precipitated by upper respiratory infection. The mucus can be dislodged by mechanical chest thumpers, and pulmonary infection can be prevented by antibiotics, so with treatment, cystic fibrosis patients can live to adulthood.

Albinism (Figure 2-16) is a rare condition that is inherited in a Mendelian manner as a recessive phenotype in many animals, including humans. The striking "white" phenotype is caused by the inability of the body to make melanin, the pigment that is responsible for most of the black and brown coloration of animals. In humans, such coloration is most evident in hair, skin and retina, and its absence in albinos (who have the homozygous recessive genotype *aa*) leads to white hair, white skin, and eye pupils that are pink because of the unmasking of the red hemoglobin pigment in blood vessels in the retina.

Figure 2-16 Albinism in an Indian man. The phenotype is caused by homozygosity for a recessive allele, say *aa*. The dominant allele *A* determines one step in the chemical synthesis of the dark pigment melanin in the cells of skin, hair, and eye retinas. In *aa* individuals this step is nonfunctional, and the synthesis of melanin is blocked. (Joe McDonald/Visuals Unlimited)

Message In pedigrees, simple Mendelian inheritance of a recessive disorder is revealed by the appearance of the phenotype in the male and female progeny of unaffected individuals.

What about disorders inherited as dominants? Here the normal allele is recessive, and the abnormal allele is dominant. It might seem paradoxical that a rare disorder can be dominant, but remember that dominance and recessiveness are simply properties of how alleles act, and are not defined in terms of predominance in the population. A good example of a rare dominant phenotype with Mendelian inheritance is achondroplasia, a type of dwarfism (see Figure 2-17). Here, people with normal stature are genotypically *dd*, and the dwarf phenotype in principle could be *Dd* or *DD*. However, it is believed that the two "doses" of the *D* allele in *DD* individuals produce such a severe effect that this is a lethal genotype. If true, all achondroplastics are heterozygotes.

In pedigree analysis, the major clues for identifying a dominant disorder with Mendelian inheritance are that the phenotype tends to appear in every generation of the pedigree and that affected fathers and mothers transmit the phenotype to both sons and daughters. Again, the representation of both sexes among the affected offspring rules out the sex-linked inheritance mentioned in our discussion of recessive disorders. The phenotype appears in every generation because generally the abnormal allele carried by an individual must have come from a parent in the previous generation. Abnormal alleles can arise de novo by the process of genetic change called mutation. This is relatively rare, but must be kept in mind as a possibility. A typical pedigree for a dominant disorder is shown in Figure 2-18. Once again, notice that Mendelian ratios are not necessarily observed in families. As with recessive disorders, individuals bearing one copy of the rare allele (*Aa*) are much more common than those bearing two copies (*AA*), so most affected people are heterozygotes, and virtually all matings involving dominant disorders are $Aa \times aa$. Therefore, when the progeny of such matings are totaled, a 1:1 ratio is expected of unaffected (*aa*) to affected individuals (*Aa*).

Figure 2-17 The human achondroplasia phenotype, illustrated by a family of five sisters and two brothers. The phenotype is determined by a dominant allele, which we can call *D*, that interferes with bone growth during development. Most members of the human population can be represented as *dd* in regard to this gene. This photograph was taken upon the arrival of the family in Israel after the end of the Second World War. (UPI/Bettmann News Photos)

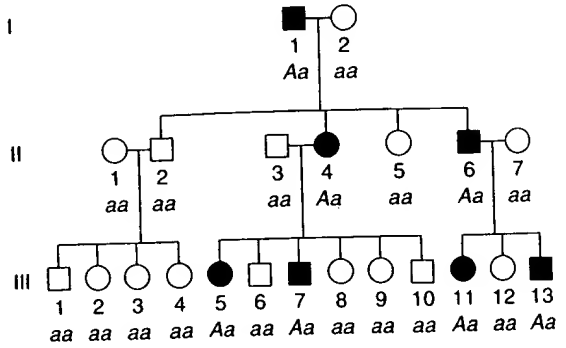


Figure 2-18 Pedigree of a dominant phenotype determined by a dominant allele *A*. In this pedigree, all of the genotypes have been deduced.

Huntington's disease is an example of a disease inherited in a Mendelian manner as a dominant phenotype. The phenotype is one of neural degeneration, leading to convulsions and premature death. However, it is a late-onset disease, the symptoms generally not appearing until after the individual has begun to have children (see Figure 2-19). Each child of a carrier of the abnormal allele stands a 50 percent chance of inheriting the allele and the associated disease. This tragic pattern has led to a great effort to find ways of identifying individuals who carry the abnormal allele before they experience the onset of the disease. The application of molecular techniques has resulted in a promising screening procedure.

Some other rare dominant conditions are polydactyly (extra digits) and brachydactyly (short digits), shown in Figure 2-20, and piebald spotting, shown in Figure 2-21.

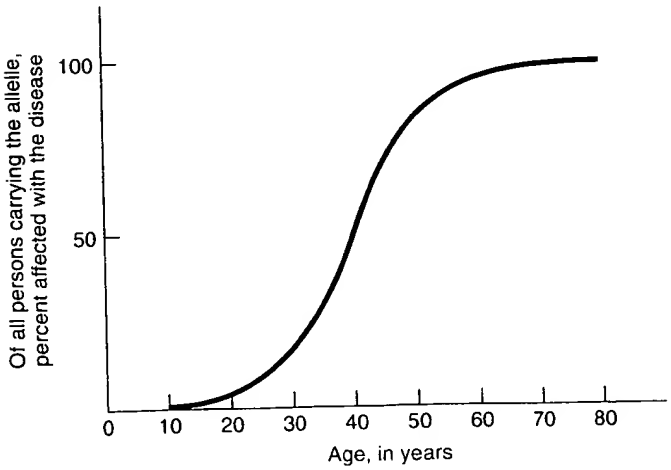
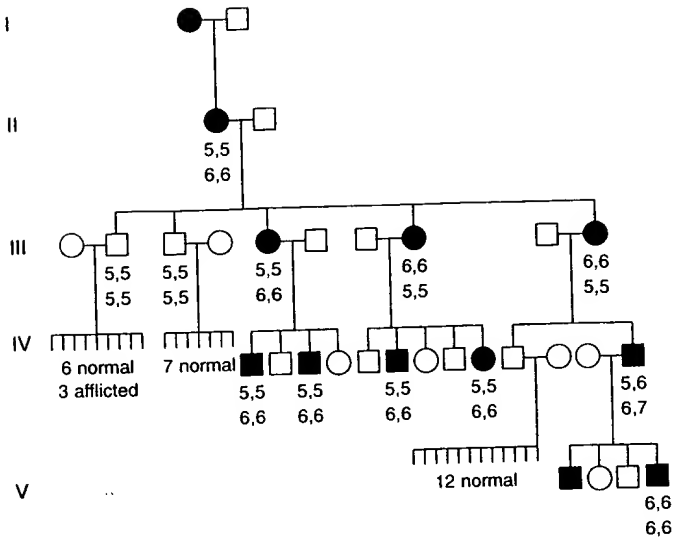


Figure 2-19 The age of onset of Huntington's disease. The graph shows that persons carrying the allele generally do not express the disease until after child-bearing age.



(a)

Figure 2-20a *(see legend on facing page.)*

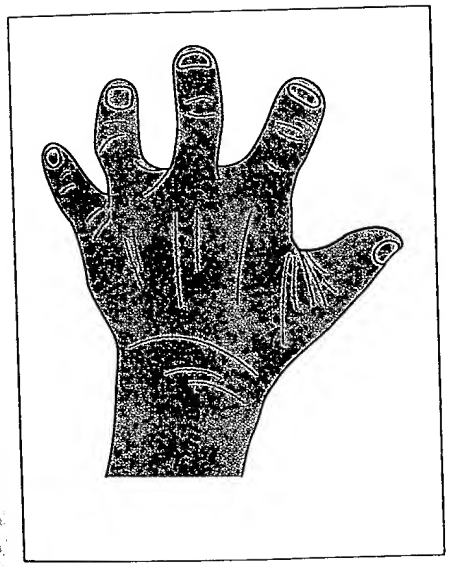
Message Pedigrees of Mendelian dominant disorders show affected males and females in each generation and also show affected men and women transmitting the condition to equal proportions of their sons and daughters.

Normal Phenotypic Variants

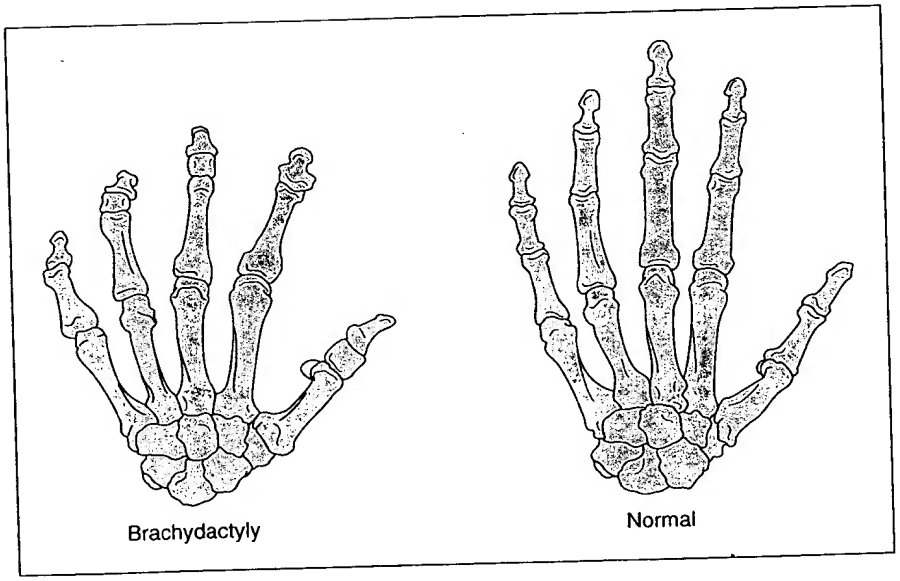
All populations of plants and animals show variation, and human populations are no exception. Because isolated populations diverge genetically, some human variation corresponds to ethnic differences. We all know, however, that even within ethnic groups there is a vast amount of variation, the type of variation that allows us to distinguish each other. Normal variation is of two types, continuous and discontinuous.

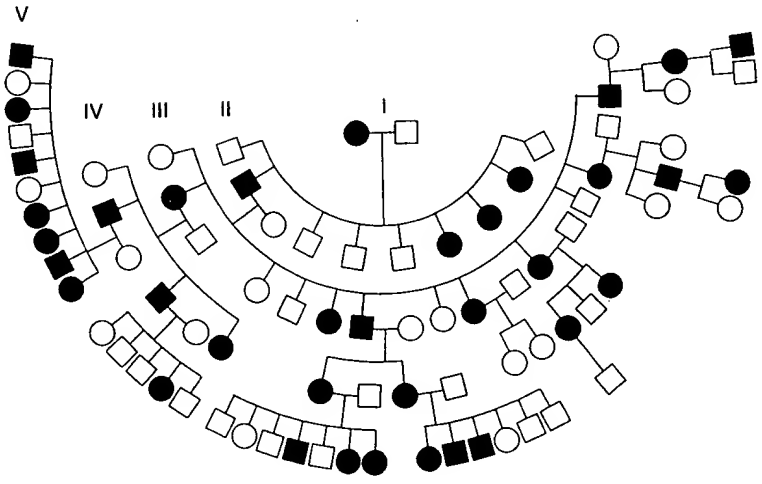
Continuous variation is that shown by measurable characters such as height or weight, in which the phenotypes are arranged along a continuous spectrum. Continuous variation requires special techniques of analysis that we cover in Chapter 24. Discontinuously varying characters, on the other hand, can be classified into distinct phenotypes, such as those analyzed by Mendel. In humans there are many examples of discontinuous characters, such as brown versus blue eyes, dark versus blonde hair, chin dimples versus none, widow's peak versus none, attached versus free earlobes, and so on. Before we look at a sample pedigree, let's consider a useful new term. A set of two or more common, alternative, normal phenotypes is called a **polymorphism,** a word derived from the Greek for "many forms." The alternative phenotypes are called morphs. Primarily we will be concerned with sets of two contrasting morphs, the simplest type of polymorphism, called dimorphism. So using one

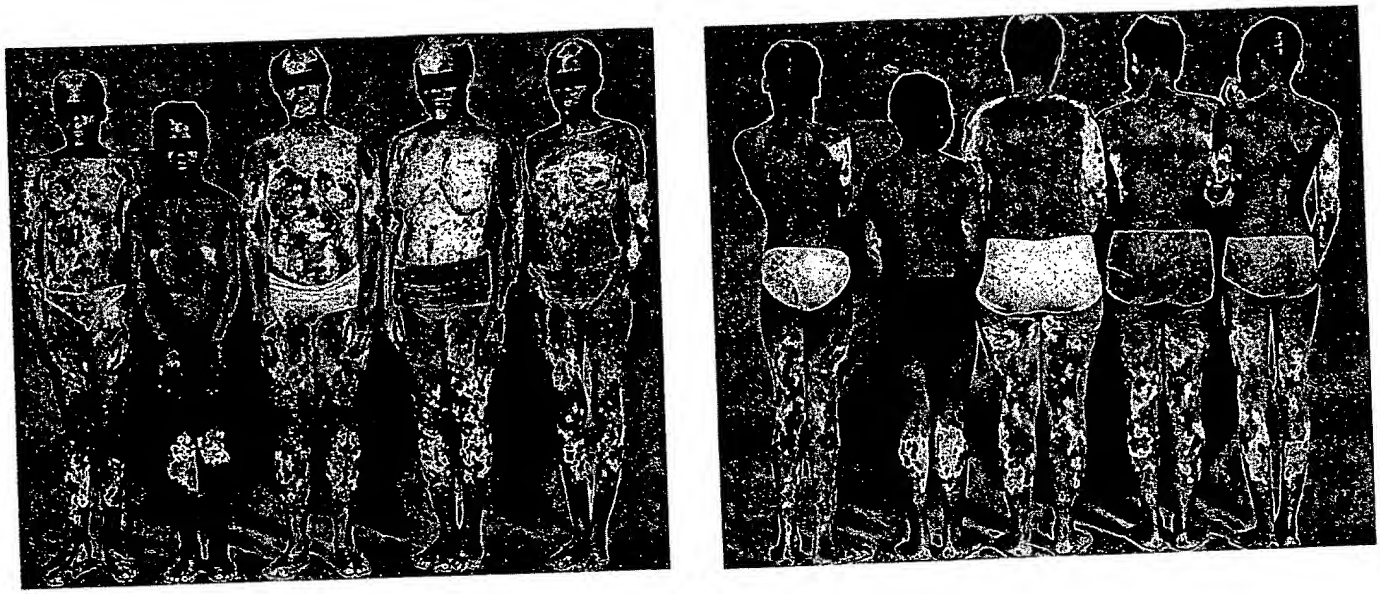
Figure 2-20 Some rare dominant phenotypes of the human hand. (a) *(left)* Polydactyly, a dominant phenotype characterized by extra fingers and/or toes, determined by an allele *P*. The numbers in the accompanying pedigree *(facing page)* show the number of fingers in the upper lines, and toes in the lower. (Note the variation in expression of the *P* allele, a topic we will cover specifically in Chapter 4.) (b) *(below)* Brachydactyly, a dominant phenotype of short fingers, determined by an allele *B*. Note the very short terminal bones in the fingers compared with those in a normal hand. The pedigree for a family with brachydactyly shows a typical inheritance pattern for a rare dominant condition. All affected individuals are *Bb* and unaffected individuals are *bb*. (a, photo: Lester Bergman & Assoc.; b, based on C. Stern, *Principles of Human Genetics,* 3d. ed., copyright 1973 by W. H. Freeman and Company.)



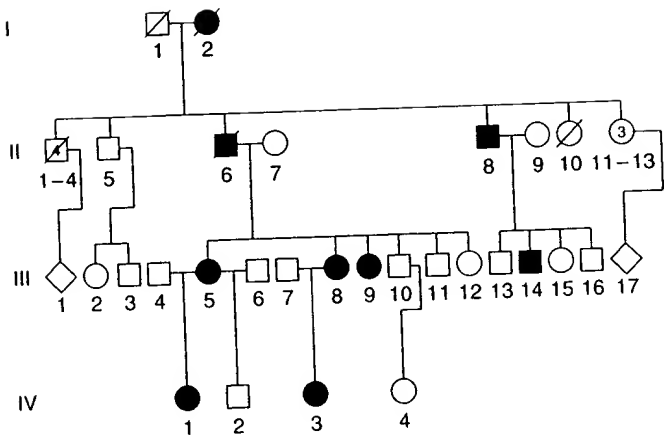
(b)



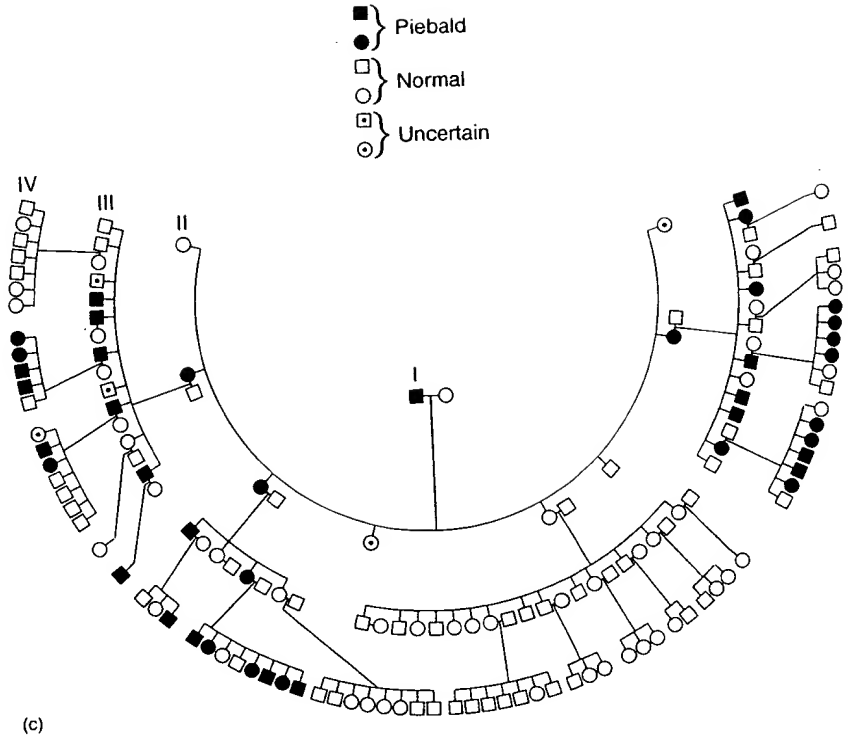




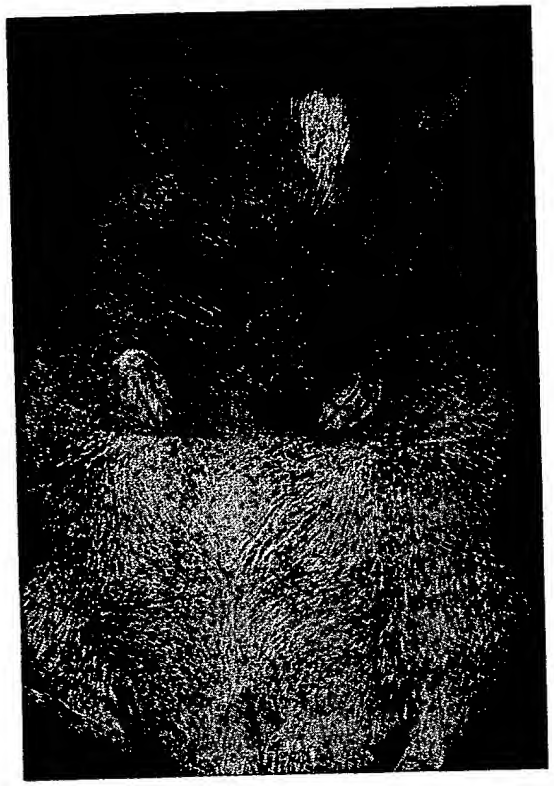
(a)


I
1
2
II
1–4
5
6
7
8
9
10
11–13
3
III
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
IV
1
2
3
4

(b)


Piebald
Normal
Uncertain
IV
III
II
I

(c)



(d)

Figure 2-21 Piebald spotting, a rare dominant human phenotype. Although the phenotype is encountered sporadically in all races, the patterns show up best in those with dark skin. (a) The photographs show front and back views of affected individuals IV-1, IV-3, III-5, III-8, and III-9 from the family pedigree shown in (b). Notice the variation between family members in expression of the piebald gene. A larger pedigree of a Norwegian family is shown in (c). It is believed that the patterns are caused by the dominant allele interfering with the migration of melanocytes (melanin-producing cells) from the dorsal to the ventral surface during development. The white forehead blaze is particularly characteristic and is often accompanied by a white forelock in the hair. The same basic condition is known in mice, and again the melanocytes fail to cover the top of the head and the ventral surface (d). Piebaldism is not a form of albinism; the cells in the light patches have the genetic potential to make melanin, but since they are not melanocytes they are not developmentally programmed to do so. In true albinism, the cells lack the potential to make melanin. (The DNA of the piebald allele has recently been characterized as an allele of c-*kit*, a type of gene called a proto-oncogene, to be discussed in Chapter 7.) (a, b from I. Winship, K. Young, R. Martell, R. Ramesar, D. Curtis, and P. Beighton, "Piebaldism: An Autonomous Autosomal Dominant Entity," *Clinical Genetics* 39, 1991, 330; c from C. Stern, *Principles of Human Genetics,* 3d ed., Copyright 1973 by W. H. Freeman and Company; d provided by R. A. Fleischman, University of Texas, Southwestern Medical Center, Dallas—also see R. A. Fleischman, D. L. Saltman, V. Stastny, and S. Zneimer, "Deletion of the c-*kit* Protooncogene in the Human Developmental Defect Piebald Trait," *Proceedings of the National Academy of Sciences USA* 88, 1991, 10885.)

of the above examples we would say that earlobes are dimorphic, with attached and free as the two major morphs. The morphs of a polymorphism are often determined by the alleles of one gene, inherited in the simple Mendelian manner described in this chapter.

The interpretation of pedigrees for polymorphisms is somewhat different, because by definition the morphs are common. Let's look at a pedigree for an interesting human dimorphism. Most human populations are dimorphic for the ability to taste the chemical phenylthiocarbamide (PTC). That is, people can either detect it as a foul, bitter taste, or—to the great surprise and disbelief of tasters—cannot taste it at all. From the pedigree in Figure 2-22, we can see that two tasters sometimes produce nontaster children. This makes it clear that the allele that confers the ability to taste is dominant and that the allele for nontasting is recessive. Notice that almost all persons who marry into this family carry the recessive allele either in heterozygous or homozygous condition. Such a pedigree thus differs from those of rare recessive abnormalities for which it is conventional to assume that all who marry into a family are homozygous normal. As both PTC alleles are common, it is not surprising that all but one of the family members in this pedigree married carriers of the recessive allele.

Polymorphism is an interesting genetic phenomenon. Population geneticists have been surprised at how much polymorphism there is in natural populations of plants and animals generally. Furthermore, even though

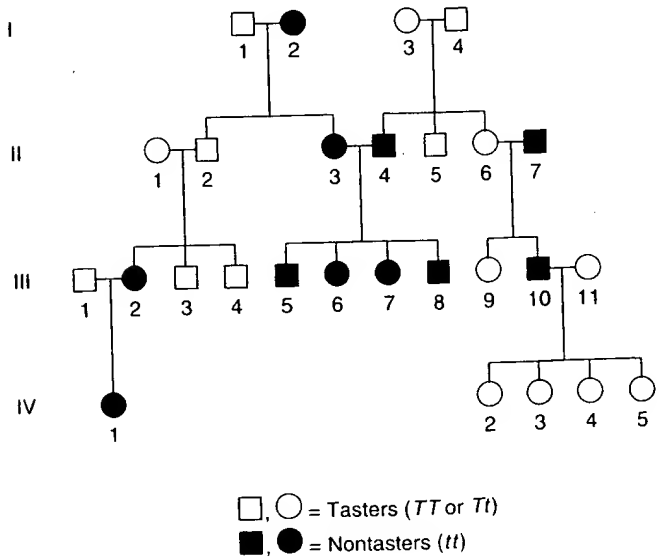
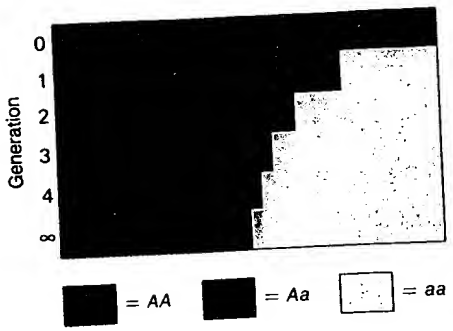


Figure 2-22 Pedigree for the ability to taste the chemical PTC.

Generation	AA	Aa	aa
0	0	100	0
1	25	50	25
2	37.5	25	37.5
3	43.75	12.5	43.75
4	46.875	6.25	46.875
∞	50	0	50

(a)



(b)

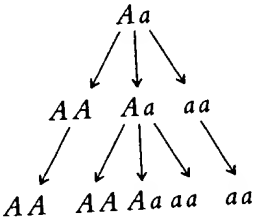
Figure 2-23 Repeated generations of selfing increase the proportion of homozygotes. (a) The percentages of the three genotypes are shown through several generations, assuming that all individuals in generation 0 are *Aa* and that all individuals reproduce at the same rate. At each generation, *AA* and *aa* breed true, but *Aa* individuals produce *Aa*, *AA*, and *aa* progeny in a 2:1:1 ratio. (b) A graphic depiction of how the proportions of genotypes change.

the genetics of polymorphisms is straightforward, there are very few polymorphisms for which there is satisfactory explanation for the coexistence of the morphs. But polymorphism is rampant at every level of genetic analysis, even down to the DNA level, and indeed polymorphisms observed at the DNA level have been invaluable as landmarks to help geneticists find their way around the chromosomes of complex organisms.

Message Populations of plants and animals (including humans) are highly polymorphic. Contrasting morphs are generally determined by alleles inherited in a simple Mendelian manner.

Simple Mendelian Genetics in Agriculture

Plant breeding methods used by Neolithic farmers were probably the same as those used until the discovery of Mendelian genetics. Basically, the approach was to select superior phenotypes that often appear as rare variants in natural populations and then propagate these over the years. Particularly desirable were pure lines of these favorable phenotypes, because such lines produced constant results over generations of planting. Without the knowledge of Mendelian genetics, how is it possible to develop pure lines? It so happens that self-pollinating plants, such as many crop plants, naturally tend to be homozygous, because selfing reduces heterozygosity in the population as is apparent from considering the genotypes of parents and the genotypes possible among their progeny:



This is illustrated in more detail in Figure 2-23. Thus, pure lines have developed automatically over the years.

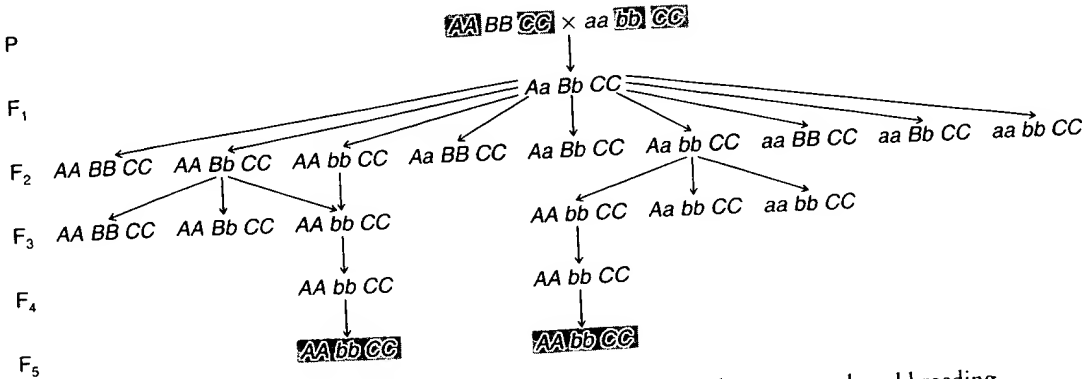


Figure 2-24 The basic technique of plant breeding. A hybrid is generated, and breeding stock is selected in subsequent generations to obtain improved pure lines. For simplicity, we assume here that the parental lines differ in only two genes (*A* and *B*). In this example, the *AA bb CC* genotype is the one desired; that is, tests on the F_2 generation show this to be the most desirable phenotype.

However, these unsophisticated breeding practices suffered from a major problem: the breeder was forced to rely on favorable natural combinations of genes. With the advent of Mendelian genetics, it became evident that favorable qualities from different lines could be combined through hybridization and subsequent gene reassortment. This procedure forms the basis of modern plant breeding.

The breeding of plants to produce new and improved genotypes works in basically the following way. For naturally self-pollinating plants, such as rice or wheat, two pure lines (each of different favorable genotype) are hybridized by manual cross-pollination and an F_1 is developed. The F_1 is then allowed to self, and its heterozygous pairs of alleles assort to produce many different genotypes, some of which represent desirable new combinations of the parental genes. A small proportion of these new genotypes will be pure-breeding already; in those that are not, several generations of selfing will produce homozygosity of the relevant genes. Figure 2-24 summarizes this method. An example of the importance of such methodology is in the breeding of highly productive, dwarf lines of rice, one of the world's staple foods. The breeding program is shown in Figure 2-25.

An example of genetic improvement in a species more familiar to most is the tomato. Anyone who has read a recent seed catalog will be familiar with the abbreviations V, F, and N next to a listed tomato variety. These represent, respectively, resistance to the pathogens *Verticillium, Fusarium,* and nematodes—resistance that has been bred into domestic tomatoes, typically from wild relatives. A useful tomato phenotype is determinate growth. The tips of the branches of most tomato species can keep growing (indeterminate growth), but in the determinate lines most branches end in a growth-arresting inflorescence. Determinate plants are bushier and more compact, and they do not need as much staking (Figure 2-26). Determinate growth is caused by a recessive allele *sp* (self-pruning), which has been crossed into modern varieties. The symbol *sp* represents a single gene; many gene symbols are more than one letter. A useful allele of another gene is *u* (uniform ripening); this allele eliminates the green patch or shoulder around the stem on the ripe fruit. Tomato geneticists have produced a large array of different morphological types of tomatoes, many of which have found their way into supermarkets and garden stores (Figure 2-27).

Such examples could be listed for many pages. The point is that simple Mendelian genetics, as described in this chapter, has provided agricultural plant breeding with its rationale and its modern methods. Entire complex genotypes may be constructed from an array of ancestral lines, each showing some desirable feature. When we think of genetic engineering, we think of the molecular techniques of the 1980s and 1990s, but genetic engineering for plant improvement began long ago.

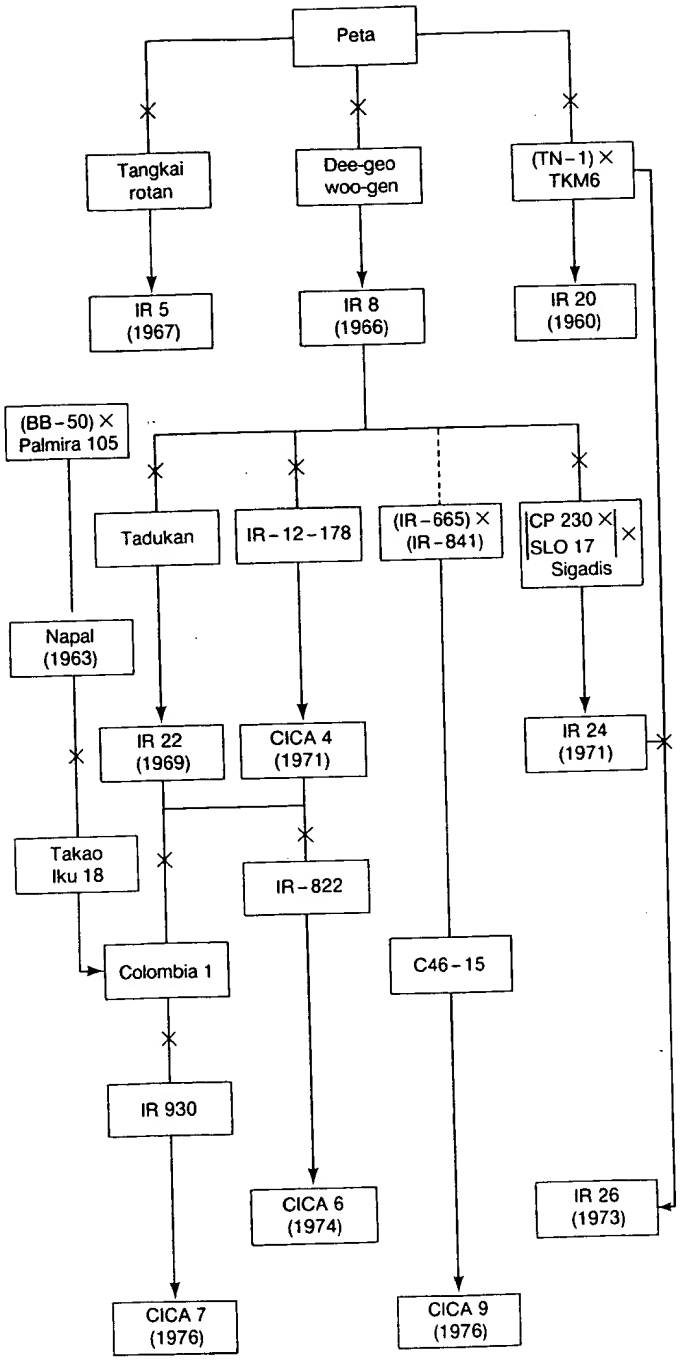


Figure 2-25 The complex pedigree of modern rice varieties. The progenitor of most modern dwarf types was IR 8, selected from a cross between the vigorous Peta (Indonesian) and the dwarf Dee-geo woo-gen (from Taiwan). Most of the other crosses represent progressive improvement of IR 8. The diagram illustrates the extensive plant breeding that produced modern crop varieties. (From Peter R. Jennings, "The Amplification of Agricultural Production." Copyright 1976 by Scientific American, Inc. All rights reserved.)

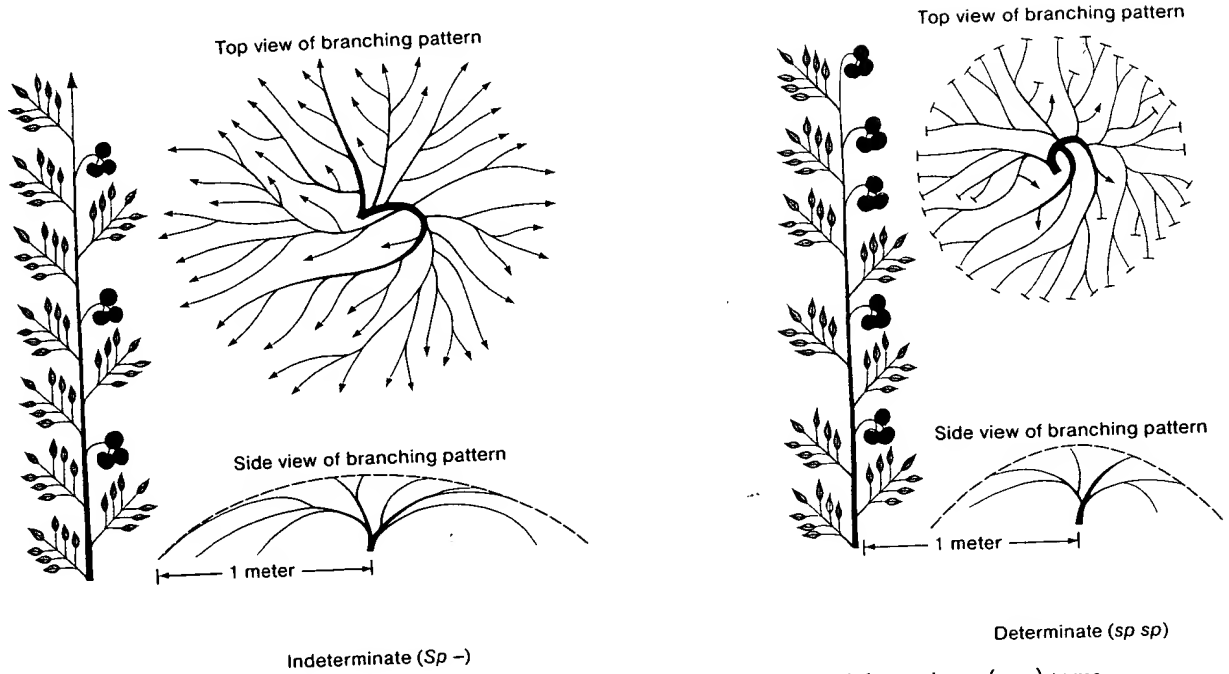


Figure 2-26 Growth characteristics of indeterminate (*Sp* –) and determinate (*spsp*) tomatoes. Arrows show growth; tomatoes indicate locations of inflorescences. Determinate has progressively fewer leaves between inflorescences, and the determinate stems end with a size-arresting inflorescence. Individual stem branches are shown, plus typical branching patterns for mature plants in top and side views. (In the top view, bars indicate inflorescences that terminate branch growth.) Note the size difference. (From Charles M. Rick, "The Tomato." Copyright 1978 by Scientific American, Inc. All rights reserved.)

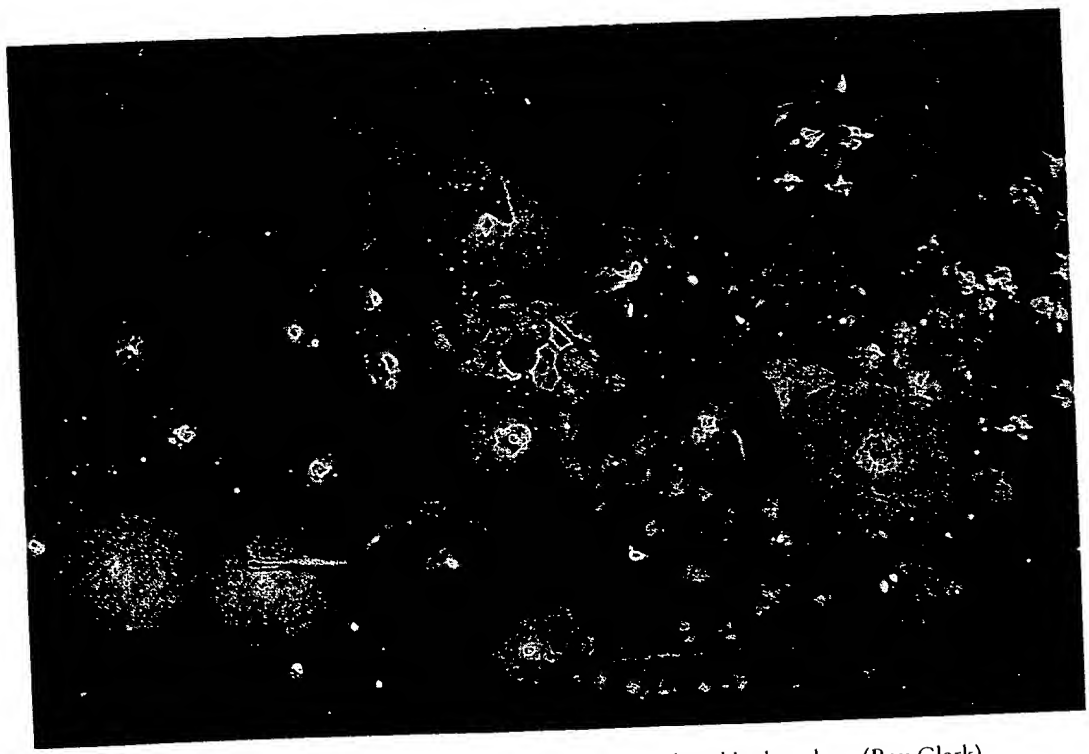


Figure 2-27 Some of the many tomato phenotypes produced by breeders. (Ray Clark)

Mendelian genetics also has provided a formal theoretical basis for animal breeding, enabling a greater efficiency than under traditional practices. More recent techniques, such as the use of frozen semen, superovulation, artificial insemination, test-tube fertilization, frozen embryos, and surrogate mothers in livestock species, have enabled breeders to amplify the number of offspring of a specific genotype—a number normally limited by the life span of the animal.

Variants and Genetic Dissection

Genetic analysis, as we have seen, must start with parental differences. Without variants, no genetic analysis is possible. Where do these variants—these raw materials for genetic analysis—come from? This is a question that can be answered in full only in later chapters. Briefly, most of the variants like the ones used by Mendel (and by ancient and modern breeders of plants and animals) arise spontaneously in nature or in the breeders' populations without the deliberate action of geneticists.

Let us emphasize again that variants can range from rare to common. Some rare variants are abnormal. Undoubtedly in a natural setting many of them would be weeded out by natural selection, but they can be kept alive by nurture so that the alleles responsible can be studied. On the other hand, for many genes there are two or more common alleles in a population, resulting in genetic polymorphism—the coexistence of genetically determined variant phenotypes in a population. Although the reason for the existence of polymorphisms is usually not easy to discover, for the geneticist they are a useful source of variant alleles for study.

We have seen that the genetic analysis of variants can identify a particular gene that is important for a biological process. This central aspect of modern genetics is called **genetic dissection.** Mendel was the first genetic surgeon. Using genetic analysis, he was able to identify and distinguish among the several components of the hereditary process in a way as convincing as if he had microdissected those components. The fact that the genes he was using were for pea shape, pea color, and so on, was largely irrelevant. Those genes were being used simply as **genetic markers,** which enabled Mendel to trace the hereditary processes of segregation and assortment. A genetic marker is a variant allele that is used to label a biological structure or process throughout the course of an experiment. It is almost as though Mendel were able to "paint" two alleles two different colors and send them through a cross to see how they behaved! Genetic markers are now routinely used in genetics and in all of biology to study all sorts of processes that the marker genes themselves do not directly affect. Markers are often morphological, as in the pea example. But molecular variants (DNA and proteins) are increasingly used as markers too.

Probably without realizing it, Mendel had also invented another aspect of genetic dissection in which the precise genes used *were* important. In this type of analysis, genetic variants are used in such a way that by studying the variant gene function, we can make inferences about the "normal" operation of the gene and the process it controls. Every time a gene is identified by Mendelian analysis, it identifies a component of a biological process.

Once a gene has been identified as affecting, say, petal color in peas, we call it a major gene, but what does this really mean? It is major in that it is obviously having a profound effect on the color of the petals. But can we conclude that it is *the* single most important step in the determination of petal color? The answer is no, and the reason may be seen in an analogy. If we were trying to discover how a car engine works, we might pull out various parts and observe the effect on the running of the engine. If a battery cable were disconnected, the engine would stop; this might lead us erroneously to conclude that this cable is *the* most important part of the running of the engine. Other parts are equally necessary, and their removal could also stop or seriously cripple the engine. In a similar way it can be shown (as we will see in Chapter 4) that several genes can be identified, all of which have a major and similar effect on petal coloration.

Mendel's work has withstood the test of time and has provided us with the basic groundwork for all modern genetic study. Yet his work went unrecognized and neglected for 35 years following its publication. Why? There are many possible reasons, but here we shall consider just one. Perhaps it was because biological science at that time could not provide evidence for any real physical units within cells that might correspond to Mendel's genetic particles. Chromosomes had certainly not yet been studied, meiosis had not yet been described, and even the full details of plant life cycles had not been worked out. Without this basic knowledge, it may have seemed that Mendel's ideas were mere numerology.

Message Mendel's work was the prototypical genetic analysis. As such, it is significant for the following reasons.

1. It showed how it is possible to study biological processes by using genetic markers.
2. It showed how the functions of genes themselves can be elucidated from the study of variant alleles.
3. It had far-reaching ramifications in agriculture and medicine.

In the next chapter, we focus on the physical locations of genes in cells and on the consequences of these locations.

Summary

Modern genetics is based on the concept of the gene, the fundamental unit of heredity. In his experiments with the garden pea, Mendel was the first to recognize the existence of genes. For example, by crossing a pure line of purple-flowered pea plants with a pure line of white-flowered pea plants and then selfing the F_1 generation, which was entirely purple, Mendel produced an F_2 generation of purple plants and white plants in a 3:1 ratio. In crosses such as those of pea plants bearing yellow seeds and pea plants bearing green seeds, he discovered that a 1:2:1 ratio underlies all 3:1 ratios. From these precise mathematical ratios Mendel concluded that there are hereditary determinants of a particulate nature (now known as genes). In higher plant and animal cells, genes exist in pairs. Variant forms of a gene are called alleles. Individual alleles can be either dominant or recessive.

In a cross of heterozygous yellow (Yy) plants with homozygous green (yy) plants, a 1:1 ratio of yellow to green plants was produced. From this ratio Mendel confirmed his so-called first law, which states that two members of a gene pair segregate from each other during gamete formation into equal numbers of gametes. Thus, each gamete carries only one member of each gene pair. The union of gametes to form a zygote is random as regards which allele the gametes carry.

The foregoing conclusions came from Mendel's work with monohybrid crosses. In dihybrid crosses, Mendel found 9:3:3:1 ratios in the F_2, which are really two 3:1 ratios combined at random. From these ratios Mendel inferred that alleles of the two genes in a dihybrid cross behave independently. This concept is Mendel's second law.

Although controlled crosses cannot be made in human beings, Mendelian genetics has great significance for humans. Many diseases and other exceptional conditions in humans are determined by recessive alleles inherited in a Mendelian manner; other exceptional conditions are caused by dominant alleles. In addition, Mendelian genetics is widely used in modern agriculture. By combining favorable qualities from different lines through hybridization and subsequent gene reassortment, plant and animal geneticists are able to produce new lines of superior phenotypes.

Finally, Mendel was responsible for the basic techniques of genetic dissection still in use today. One such technique is the use of genes as genetic markers to trace the hereditary processes of segregation and assortment. The other is the study of abnormal variants to discover how genes operate normally.

Chapter Integration Problem

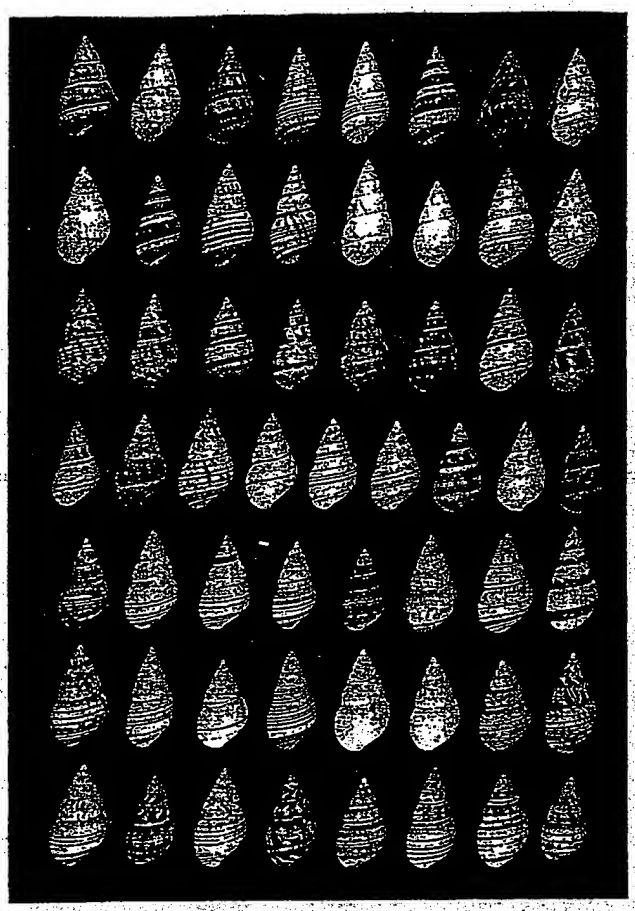
Each chapter has one solved problem in which we stress the integration of concepts from different chapters. Learning in any discipline is a linear process, moving from topic to topic in some kind of appropriate sequence. But of course the discipline itself is not linear, but a set of integrated parts that the professional sees as one whole. We hope that focusing on integration will clarify the overall structure of genetics and that the reader will not see the contents of each chapter in isolation. As we pass from chapter to chapter, the levels of understanding build on the previous ones, and the subject is assembled like the layers of an onion.

Crosses were made between two pure lines of rabbits that we can call A and B. A male from line A was mated with a female from line B, and the F_1 rabbits were subsequently intercrossed to produce an F_2. It was discovered that $\frac{3}{4}$ of the F_2 animals had white subcutaneous fat, and $\frac{1}{4}$ had yellow subcutaneous fat. Later, the F_1 was examined and was found to have white fat. Several years later, an attempt was made to repeat the experiment using the same male from line A and the same female from line B. This time, the F_1 and all the F_2 (22 animals) had white fat. The only difference between the original experiment and the repeat that seemed relevant was that in the original all the animals were fed on fresh vegetables, and in the repeat they were fed on commercial rabbit chow. Provide an explanation for the difference and a test of your idea.

Solution

The first time the experiment was done, the breeders would have been perfectly justified in proposing that a pair of alleles determine white versus yellow body fat. This is because the data clearly resemble Mendel's results in peas. White must be dominant, so we can represent the white allele as W, and the yellow allele as w. The results can then be expressed

25
Population Genetics



Shell color polymorphism in *Liguus fasciatus*. (From David Hillis, *Journal of Heredity*, July–August 1991.)

KEY CONCEPTS

The goal of population genetics is to understand the genetic composition of a population and the forces that determine and change that composition.

In any species, a great deal of genetic variation within and between populations arises from the existence of various alleles at different gene loci.

A fundamental measurement in population genetics is the frequency at which the alleles can occur at any gene locus of interest.

The frequency of a given allele in a population can be changed by recurrent mutation, selection, or migration, or by random sampling effects.

In an idealized population, in which no forces of change are acting, a randomly interbreeding population would show constant genotypic frequencies for a given locus.

Mendel's investigations of heredity — indeed all the interest in heredity in the nineteenth century — arose from two related problems: how to breed improved crops and how to understand the nature and origin of species. What is common to these problems (and differentiates them from the problems of transmission and gene action) is that they are concerned with *populations* rather than with *individuals*. Studies of gene replication, protein synthesis, development, and chromosome movement focus on processes that go on within the cells of individual organisms. But the transformation of a species, either in the natural course of evolution or by the deliberate intervention of human beings, is a change in the properties of a collectivity — of an entire population or a set of populations.

Message Population genetics relates the heritable changes in populations of organisms to the underlying individual processes of inheritance and development.

Darwin's Revolution

The modern theory of evolution is so completely identified with the name of Charles Darwin (1809–1882) that many people think that the concept of organic evolution was first proposed by Darwin, but that is certainly not the case. Most scholars had abandoned the notion of fixed species, unchanged since their origin in a grand creation of life, long before publication of Darwin's *The Origin of Species* in 1859. By that time, most biologists agreed that new species arise through some process of evolution from older species; the problem was to explain *how* this evolution could occur.

Darwin's theory of the mechanism of evolution begins with the variation that exists among organisms within a species. Individuals of one generation are qualitatively different from one another. Evolution of the species as a whole results from the differential rates of survival and reproduction of the various types, so that the relative frequencies of the types change over time. Evolution, in this view, is a sorting process. For Darwin, evolution of the group resulted from the differential survival and reproduction of individual variants *already existing* in the group — variants arising in a way unrelated to the environment.

Message Darwin proposed a new explanation to account for the accepted phenomenon of evolution. He argued that the population of a given species at a given time includes individuals of varying characteristics. The population of the next generation will contain a higher frequency of those types that most successfully survive and reproduce under the existing environmental conditions. Thus, the frequencies of various types within the species will change over time.

There is an obvious similarity between the process of evolution as Darwin described it and the process by which the plant or animal breeder improves a domestic stock. The plant breeder selects the highest-yielding plants from the current population and (as far as possible) uses them as the parents of the next generation. If the characteristics causing the higher yield are heritable, then the next generation should produce a higher yield. It was no accident that Darwin chose the term **natural selection** to describe his model of evolution through differential rates of reproduction of different variants in the population. As a model for this evolutionary process, he had in mind the selection that breeders exercise on successive generations of domestic plants and animals.

We can summarize Darwin's theory of evolution through natural selection in three principles:

1. **Principle of variation.** Among individuals within any population, there is variation in morphology, physiology, and behavior.
2. **Principle of heredity.** Offspring resemble their parents more than they resemble unrelated individuals.
3. **Principle of selection.** Some forms are more successful at surviving and reproducing than other forms in a given environment.

Clearly, a selective process can produce change in the population composition only if there are some variations to select among. If all individuals are identical, no amount of differential reproduction of individuals can affect the composition of the population. Furthermore, the variation must be in some part heritable if differential reproduction is to alter the population's genetic composition. If large animals within a population have more offspring than small ones but their offspring are no larger on average than those of small animals, then no change in population composition can occur from one generation to another. Finally, if all variant types leave, on average, the same number of offspring, then we can expect the population to remain unchanged.

Message Darwin's principles of variation, heredity, and selection must hold true if there is to be evolution by a variational mechanism.

Variation and Its Modulation

Population genetics is the translation of Darwin's three principles into precise genetic terms. As such, it deals with the description of genetic variation in populations and with the experimental and theoretical determination of how that variation changes in time and space.

Message Population genetics is the study of inherited variation and its modulation in time and space.

Observations of Variation

Population genetics necessarily deals with genotypic variation, but by definition, only phenotypic variation can be observed. The relation between phenotype and genotype varies in simplicity from character to character. At one extreme, the phenotype may be the observed DNA sequence of a stretch of the genome. In this case, the distinction between genotype and phenotype disappears, and we can say that we are, in fact, directly observing the genotype. At the other extreme lie the bulk of characters of interest to plant and animal breeders and to most evolutionists — the variations in yield, growth rate, body shape, metabolic ratio, and behavior that constitute the obvious differences between varieties and species. These characters have a very complex relation to genotype, and we must use the methods introduced in Chapter 24 to say anything at all about the genotypes. But as we have seen in Chapter 24, it is not possible to make very precise statements about the genotypic variation underlying quantitative characters. For that reason, most of the study of experimental population genetics has concentrated on characters with simple relations to the genotype, much like the characters studied by Mendel. A favorite object of study for human population geneticists, for example, has been the various human blood groups. The qualitatively distinct phenotypes of a given blood group — say, the MN group — are coded for by alternative alleles at a single locus, and the phenotypes are insensitive to environmental variations.

The study of variation, then, consists of two stages. The first is a description of the phenotypic variation. The second is a translation of these phenotypes into genetic terms and the redescription of the variation genetically. If there is a perfect one-to-one correspondence between genotype and phenotype, then these two steps merge into one, as in the case of the MN blood group. If the relation is more complex — for example, as the result of dominance, so that heterozygotes resemble homozygotes, it may be necessary to carry out experimental crosses or to observe pedigrees to translate phenotypes into genotypes. This is the case for the human ABO blood group (see page 89).

The simplest description of Mendelian variation is the frequency distribution of genotypes in a population. Table 25-1 shows the frequency distribution of the three genotypes at the MN blood-group locus in several human populations. Note that there is variation both within and between populations. More typically, instead of the frequencies of the diploid genotypes, the frequencies of the alternative alleles are used. If f_{AA}, f_{Aa}, and f_{aa} are the proportions of the three genotypes at a locus with two alleles, then the frequencies $p(A)$ and $q(a)$ of the alleles are obtained by counting alleles. Since each homozygote AA consists only of A alleles and only half the alleles of each heterozygote Aa are type A, the total frequency (p) of A alleles in the population is

$$p = f_{AA} + \tfrac{1}{2}f_{Aa} = \text{frequency of } A$$

Similarly, the frequency q of a alleles is given by

$$q = f_{aa} + \tfrac{1}{2}f_{Aa} = \text{frequency of } a$$

$$p + q = f_{AA} + f_{aa} + f_{Aa} = 1.00$$

If there are multiple alleles, then the frequency for each allele is simply the frequency of its homozygote plus half the sum of the frequencies for all the heterozygotes in which it appears. Table 25-1 shows the values of p and q for each of the MN blood-group populations.

Table 25-1. Frequencies of Genotypes for Alleles at MN Blood-Group Locus in Various Human Populations

	Genotype			Allele frequencies	
Population	*MM*	*MN*	*NN*	*p*(*M*)	*q*(*N*)
Eskimo	0.835	0.156	0.009	0.913	0.087
Australian aborigine	0.024	0.304	0.672	0.176	0.824
Egyptian	0.278	0.489	0.233	0.523	0.477
German	0.297	0.507	0.196	0.550	0.450
Chinese	0.332	0.486	0.182	0.575	0.425
Nigerian	0.301	0.495	0.204	0.548	0.452

SOURCE: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

As an extension of p, which represents the **gene frequency** or **allele frequency,** we can describe variation at more than one locus simultaneously in terms of the **gametic frequencies.** Locus *S* (the secretor factor) is closely linked to the *MN* locus in humans. Table 25-2 shows the gametic frequencies of the four gametic types (*MS*, *Ms*, *NS*, and *Ns*) in various populations. The gametic frequency is obtained by summing up all the contributions of the different heterozygotes and homozygotes to the total pool of gametes. For example, the frequency of the *MS* gamete is given by

$$g(MS) = \text{frequency of } MS/MS + \tfrac{1}{2} \text{ frequency of } MS/NS + \tfrac{1}{2} \text{ frequency } MS/Ms + \tfrac{1}{2} \text{ frequency of } MS/Ns$$

Note that the last term in the sum involves the frequency of the double heterozygote *MS* / *Ns*. There is no contribution from the other double heterozygote *Ms* / *NS*, because it produces no *MS* gametes. If there were recombination, the *Ms* / *NS* heterozygote would produce *MS* gametes at a rate proportional to the recombination frac-

Table 25-2. Frequencies of Gametic Types for MNS System in Various Human Populations

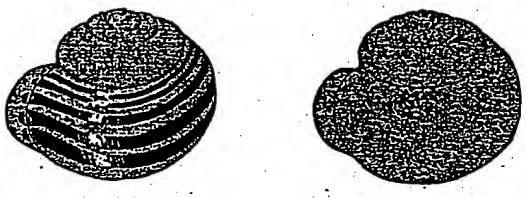
Population	Gametic type				Heterozygosity (H)	
	MS	*Ms*	*NS*	*Ns*	From gametes	From alleles
Ainu	0.024	0.381	0.247	0.348	0.672	0.438
Ugandan	0.134	0.357	0.071	0.438	0.658	0.412
Pakistan	0.177	0.405	0.127	0.291	0.704	0.455
English	0.247	0.283	0.080	0.390	0.700	0.469
Navaho	0.185	0.702	0.062	0.051	0.467	0.286

SOURCE: A. E. Mourant, *The Distribution of the Human Blood Groups.* Blackwell Scientific Pub., 1954.

tion. Thus, to give a gametic frequency description of a population at more than one locus simultaneously, we need to be able to distinguish coupling from repulsion double heterozygotes and to know the recombination fraction between the genes. For the human MNS system, there is essentially no recombination, and the different types of heterozygotes can be distinguished by pedigree analysis.

A *measure* of genetic variation (as opposed to its *description* by gene frequencies) is the amount of **heterozygosity** at a locus in a population, which is given by the total frequency of heterozygotes at a locus. If one allele is in very high frequency and all others are near zero, then there will be very little heterozygosity because, by necessity, most individuals will be homozygous for the common allele. We expect heterozygosity to be greatest when there are many alleles at a locus, all at equal frequency. In Table 25-1, the heterozygosity is simply equal to the frequency of the MN genotype in each population. When more than one locus is considered, there are two possible ways of calculating heterozygosity. First, we can average the frequency of heterozygotes at each locus separately. Alternatively, we can take the gametic frequencies, as in Table 25-2, and calculate the proportion of all individuals who carry two different gametic forms. The results of both calculations are given in Table 25-2. (See the discussion of Hardy-Weinberg equilibrium on page 750 for the calculation of heterozygosity.)

Simple Mendelian variation can be observed within and between populations of any species at various levels of phenotype, from external morphology down to the amino acid sequence of enzymes and other proteins. Indeed, with the new methods of DNA sequencing, variations in DNA sequence (such as third-position variants that are not differentially coded in amino acid sequences and even variations in nontranslated intervening sequences) have been observed. Every species of organism ever examined has revealed considerable genetic variation, or **polymorphism,** reflected at one or more levels of phenotype, either within populations or between populations, or both. Genetic variation that might be the basis for evolutionary change is ubiquitous. The tasks for population geneticists are to describe that ubiquitous variation quantitatively in terms that allow evolutionary predictions and to build a theory of evolutionary change that can use these observations in prediction.



(a) (b)

Figure 25-1 Shell patterns of the snail *Cepaea nemoralis:* (a) banded yellow; (b) unbanded pink.

It is quite impossible in this text to provide an adequate picture of the immense richness of genetic variation that exists in species. We can consider only a few examples of the different kinds of Mendelian variation to gain a superficial sense of the genetic diversity within species. Each of these examples can be multiplied many times over in other species and with other traits.

Morphological Variation. The shell of the land snail *Cepaea nemoralis* may be pink or yellow, depending on two alleles at a single locus, with pink dominant to yellow. Also, the shell may be banded or unbanded (Figure 25-1) as a result of segregation at a second linked locus, with unbanded dominant to banded. Table 25-3 shows the variation of these two loci in several European colonies of the snail. The populations also show polymorphism for the number of bands and the height of the shells, but these characters have complex genetic bases.

Examples of naturally occurring morphological variation within plant species are *Plectritis* (see Figure 1-8); *Collinsia* (blue-eyed Mary, page 46), and clover (see Figure 4-5).

Chromosomal Polymorphism. Although the karyotype is often regarded as a distinctive characteristic of a

Table 25-3. Frequencies of Snails *(Cepaea nemoralis)* with Different Shell Colors and Banding Patterns in Three French Populations

Population	Yellow		Pink	
	Banded	Unbanded	Banded	Unbanded
Guyancourt	0.440	0.040	0.337	0.183
Lonchez	0.196	0.145	0.564	0.095
Peyresourde	0.175	0.662	0.100	0.062

SOURCE: Maxime Lamotte, *Bulletin Biologique de France et Belgique,* supplement 35, 1951.

Table 25-4. Frequencies of Plants with Supernumerary Chromosomes and of Translocation Heterozygotes in a Population of *Clarkia elegans* from California

No supernumeraries or translocations	Supernumeraries	Translocations	Both translocations and supernumeraries
0.560	0.265	0.133	0.042

SOURCE: H. Lewis, *Evolution* 5, 1951, 142–157.

species, in fact, numerous species are polymorphic for chromosome number and morphology. Extra chromosomes (supernumeraries), reciprocal translocations, and inversions segregate in many populations of plants, insects, and even mammals.

Table 25-4 gives the frequencies of supernumerary chromosomes and translocation heterozygotes in a population of the plant *Clarkia elegans* from California. The "typical" species karyotype would be hard to identify.

Immunological Polymorphism. A number of loci in vertebrates code for antigenic specificities such as the ABO blood types. Over 40 different specificities on human red cells are known, and several hundred are known in cattle. Another major polymorphism in humans is the HLA system of cellular antigens, which are implicated in tissue graft compatibility (Chapter 17). Table 25-5 gives the allelic frequencies for the ABO blood-group locus in some very different human populations. The polymorphism for the HLA system is vastly greater. There appear to be two main loci, each with five distinguishable alleles. Thus, there are $5^2 = 25$ different possible gametic types, making 25 different homozygous forms and $(25)(24)/2 = 300$ different heterozygotes. All genotypes are not phenotypically distinguishable, however, so only 121 phenotypic classes can be seen. L. L. Cavalli-Sforza and W. F. Bodmer report that, in a sample of only 100 Europeans, 53 of the 121 possible phenotypes were actually observed!

Protein Polymorphism. In recent years, studies of genetic polymorphism have been carried down to the level of the polypeptides coded by the structural genes themselves. If there is a nonredundant codon change in a structural gene (say, GGU to GAU), this will result in an amino acid substitution in the polypeptide produced at translation (in this case, glycine to aspartic acid). If a specific protein could be purified and sequenced from separate individuals, then it would be possible to detect genetic variation in a population at this level. In practice, this is tedious for large organisms and impossible for small ones unless a large mass of protein can be produced from a homozygous line.

There is, however, a practical substitute for sequencing that makes use of the change in the physical properties of a protein when an amino acid is substituted. Five amino acids (glutamic acid, aspartic acid, arginine, lysine, and histidine) have ionizable side chains that give a protein a characteristic net charge, depending on the pH of the surrounding medium. Amino acid substitutions may directly replace one of these charged amino acids, or a noncharged substitution near one of them in the polypeptide chain may affect the degree of ionization of the charged amino acid, or a substitution at the joining between two α helices may cause a slight shift in the three-dimensional packing of the folded polypeptide. In all these cases, the net charge on the polypeptide will be altered.

To detect the change in net charge, protein can be subjected to the method of gel electrophoresis. Figure 25-2 shows the outcome of such an electrophoretic sepa-

Table 25-5. Frequencies of the Alleles I^A, I^B, and i at the ABO Blood-Group Locus in Various Human Populations

Population	I^A	I^B	i
Eskimo	0.333	0.026	0.641
Sioux	0.035	0.010	0.955
Belgian	0.257	0.058	0.684
Japanese	0.279	0.172	0.549
Pygmy	0.227	0.219	0.554

SOURCE: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

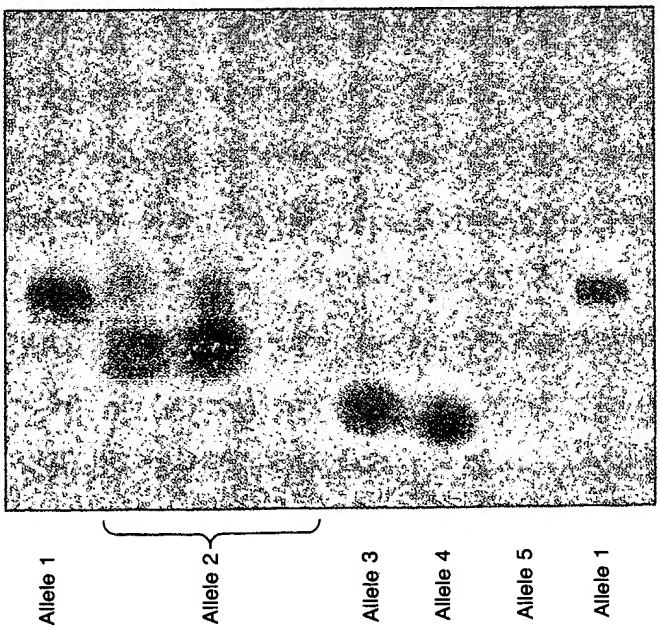


Figure 25-2 Electrophoretic gel showing homozygotes for five different alleles at the *esterase-5* locus in *Drosophila pseudoobscura*. Repeated samples of the same allele are identical, but there are repeatable differences between alleles.

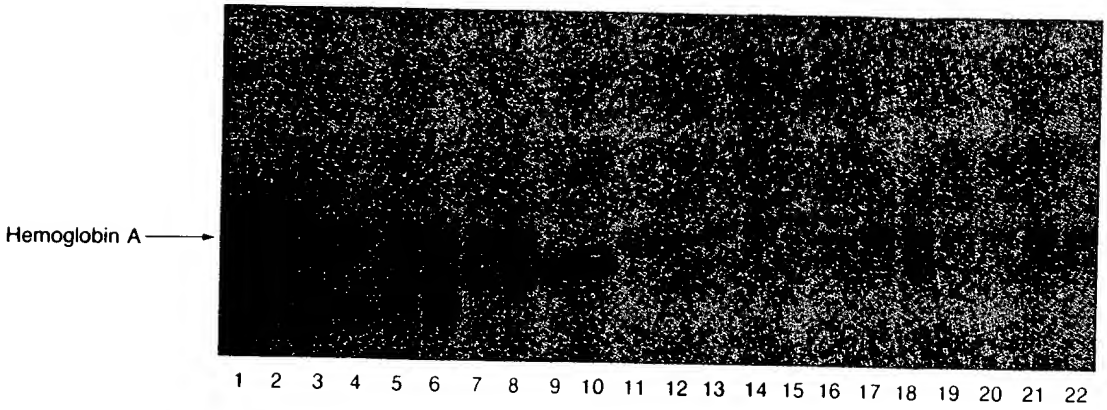


Figure 25-3 Electrophoretic gel showing heterozygotes of normal hemoglobin A and a number of variant hemoglobin alleles. One of the dark-staining bands is marked as hemoglobin A. The second dark-staining band in each track (seen most clearly in tracks 3 and 4) represents the second protein derived from the second allele of the heterozygote. Hemoglobin A is missing from tracks 9 and 10. (The dark-staining band with the same electrophoretic mobility in all tracks toward the top of the gel represents a protein other than hemoglobin.)

ration of variants of an esterase enzyme in *D. pseudoobscura*, where each track is the protein from a different individual. Figure 25-3 shows a similar gel for different variant human hemoglobins. In this case, most individuals are heterozygous for the variant and normal hemoglobin A. Table 25-6 shows the frequencies of different alleles for three enzyme-coding loci in *D. pseudoobscura* in several populations: a nearly monomorphic locus (malic dehydrogenase), a moderately polymorphic locus (α-amylase), and a highly polymorphic locus (xanthine dehydrogenase).

The technique of gel electrophoresis (or sequencing) differs fundamentally from other methods of genetic analysis in allowing the study of loci that are not segregating, because the presence of a polypeptide is prima facie evidence of a structural gene. Thus, it has been possible to ask what proportion of all structural genes in the genome of a species is polymorphic and what the average heterozygosity is in a population. Very large numbers of species have been sampled by this method, including bacteria, fungi, higher plants, vertebrates, and invertebrates. The results are remarkably consistent over species. About one-third of structural-gene loci are polymorphic, and the average heterozygosity in a population over all loci sampled is about 10 percent. This means that scanning the genome in virtually any species would show that about 1 in every 10 loci is in heterozygous condition and that about one-third of all loci have two or more

Table 25-6. Frequencies of Various Alleles at Three Enzyme-Coding Loci in Four Populations of *Drosophila pseudoobscura*

Locus (enzyme encoded)	Allele	Population			
		Berkeley	Mesa Verde	Austin	Bogotá
Malic dehydrogenase	*A*	0.969	0.948	0.957	1.00
	B	0.031	0.052	0.043	0.00
α-Amylase	*A*	0.030	0.000	0.000	0.00
	B	0.290	0.211	0.125	1.00
	C	0.680	0.789	0.875	0.00
Xanthine dehydrogenase	*A*	0.053	0.016	0.018	0.00
	B	0.074	0.073	0.036	0.00
	C	0.263	0.300	0.232	0.00
	D	0.600	0.581	0.661	1.00
	E	0.010	0.030	0.053	0.00

SOURCE: R. C. Lewontin, *The Genetic Basis of Evolutionary Change.* Columbia University Press, 1974.

alleles segregating in any population. This represents an immense potential of variation for evolution. The disadvantage of the electrophoretic technique is that it detects variation only in structural genes. If most of the evolution of shape, physiology, and behavior rests on changes in regulatory genetic elements, then the observed variation in structural genes would be beside the point.

DNA Sequence Polymorphism

Recent advances in the techniques of DNA analysis have made it possible to examine variation among individuals and between species in their DNA sequences. There are two levels at which such studies can be done. Studying variation in the sites recognized by restriction enzymes provides a coarse view of base-pair variation. At a finer level, methods of DNA sequencing allow variation to be observed base pair by base pair.

Restriction Site Variation. A six-cutting restriction enzyme will recognize an appropriate six-base sequence approximately once every $4^6 = 4096$ base pairs along a DNA molecule (determined from the probability that a specific base [of which there are four] will be found at each of the six positions). If there is polymorphism in the population for one of the six bases at the recognition site, then there will be a restriction fragment length polymorphism (RFLP) in the population, because in one variant the enzyme will recognize and cut the DNA, while in the other variant it will not (see pages 456–457). A panel of, say, eight enzymes will then sample every $4096/8 \cong 500$ base pairs for such polymorphisms. Of course, when one is found, we do not know which of the six base pairs at the recognition site is polymorphic. Using four-cutting enzymes, there is a recognition site every $4^4 = 256$ base pairs, so a panel of eight different enzymes can sample about once every 32 base pairs along the enzyme. In addition to single base-pair changes that destroy restriction enzyme recognition sites, there are insertions and deletions of stretches of DNA that also cause restriction fragment lengths to vary.

Extensive samples have been made for several regions of the genome in a number of species of *Drosophila* using both four-cutting and six-cutting enzymes. The result of one such study of the xanthine dehydrogenase gene in *Drosophila pseudoobscura* is shown in Figure 25-4. The figure shows, symbolically, the restriction pattern of 53 chromosomes (haplotypes) sampled from nature, po-

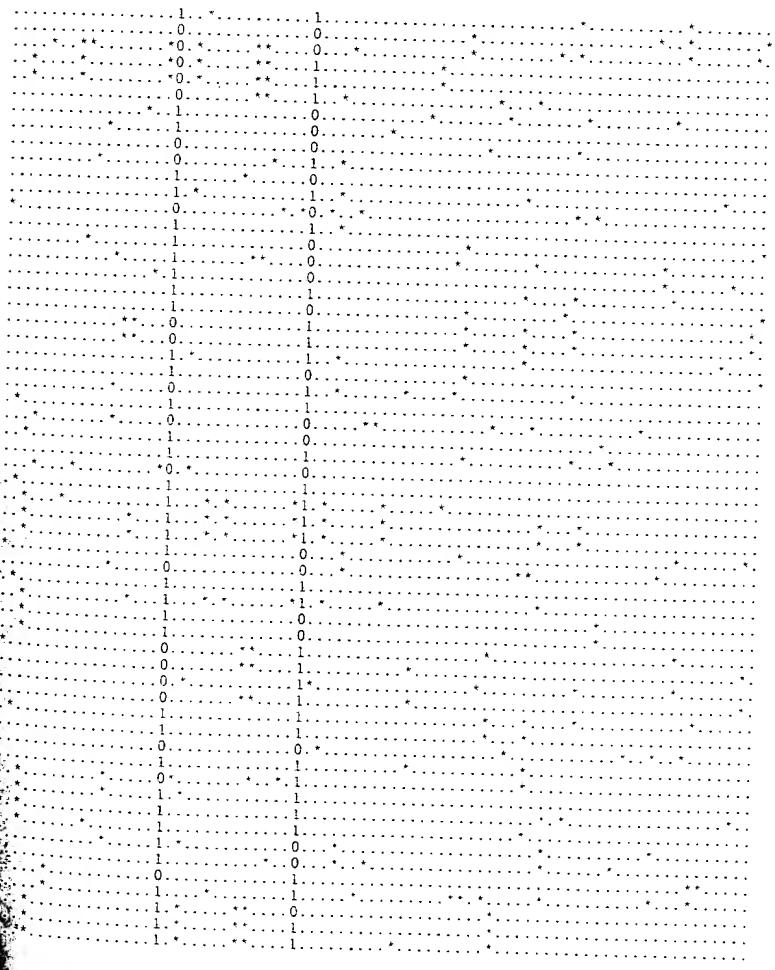


Figure 25-4 The result of a four-cutter survey of 53 chromosomes, probed for the xanthine dehydrogenase gene in *Drosophila pseudoobscura*. Each line is a chromosome (haplotype) sampled from a natural population. Each position along the line is a polymorphic restriction rate along the 4.5-kb sequence studied. Where an asterisk appears, the haplotype differs from the majority, either cutting where most haplotypes are not cut, or not cutting where most haplotypes are cut. At two sites there is no clear majority type, so a 0 or 1 is used to show whether the site is absent or present.

lymorphic for 78 restriction sites along a sequence 4.5 kb in length. Among the 53 haplotypes there are 48 different ones. (The reader is challenged to find the identical pairs.) Clearly there is an immense amount of nucleotide variation at the xanthine dehydrogenase locus in nature.

Twenty studies of different regions of the X chromosome and the two large autosomes of *Drosophila melanogaster* by restriction enzymes have found between 0.1 and 1.0 percent heterozygosity per nucleotide site, with an average of 0.4 percent. A study of the very small fourth chromosome, however, found no polymorphism at all.

Tandem Repeats. Another form of DNA sequence variation that can be revealed by restriction fragment surveys arises from the occurrence of multiply repeated DNA sequences. In the human genome, there are a variety of different short DNA sequences dispersed through the genome, each one of which is multiply repeated in a tandem row. The number of repeats may vary from a dozen to more than 100 in different individual genomes. Such sequences are known as **variable number tandem repeats (VNTR).** If the DNA is cut by restriction enzymes that flank either side of such a tandem array, a fragment will be produced whose size is proportional to the number of repeated elements. The different-sized fragments will migrate at different rates in an electrophoretic gel. Unfortunately, the individual elements are too short to allow distinguishing between, say, 64 and 68 repeats, but size classes (*bins*) can be established, and a population can be assayed for the frequencies of the different classes. Table 25-7 shows the data for two different VNTRs sampled in two American Indian groups from Brazil. In one case, D14S1, the Karitiana are nearly homozygous while the Surui are very variable, while in the other, D14S13, both populations are variable but with different frequency patterns.

Table 25-7. Size Class Frequencies for Two Different VNTR Sequences, D14S1 and D14S13, in the Karitiana and Surui of Brazil.

Size class in kb	D14S1		D14S13	
	Karitiana	Surui	Karitiana	Surui
3–4	105	4	0	0
4–5	0	3	3	14
5–6	0	11	1	4
6–7	0	2	1	2
7–8	0	1	1	2
8–9	3	3	8	16
9–10	0	11	28	9
10–11	0	2	22	0
11–12	0	4	18	8
12–13	0	0	13	18
13–14	0	0	13	3
>14	0	0	0	2
	108	78	108	78

SOURCE: Data of J. Kidd and K. Kidd. *American Journal of Physical Anthropology* 81, 1992, 249.

Complete Sequence Variation. Studies of variation at the level of single base pairs by DNA sequencing can provide information of two kinds. First, by translating the sequences of coding regions obtained from different individuals in a population or from different species, the exact amino acid sequence differences can be determined. Electrophoretic studies show only that there is variation in amino acid sequences, but cannot identify how many or which amino acids differ between individuals. So, when DNA sequences were obtained for the various electrophoretic variants of esterase-5 in *Drosophila pseudoobscura* (see Figure 25-2), it was found that electrophoretic classes differ from each other by an average of 8 amino acids, and the 20 different kinds of amino acids were involved in polymorphisms at about the frequency that they were represented in the protein. Such studies also show that different regions of the same protein have different amounts of polymorphism. For the esterase-5 protein, consisting of 545 amino acids, 7 percent of amino acid positions are polymorphic, but the last amino acids at the carboxyl terminus of the protein are totally invariant between individuals.

Second, DNA base-pair variation can also be studied for those base pairs that do not determine or change the protein sequence. This includes DNA in introns, in 5′-flanking sequences that may be regulatory, in nontranscribed DNA 3′ to the gene, and in those nucleotide positions within codons (usually third positions) whose variation does not result in amino acid substitutions. Within coding sequences, these so-called silent or synonymous base-pair polymorphisms are much more common than changes that result in amino acid polymorphism, presumably because many amino acid changes interfere with normal function of the protein and are eliminated by natural selection. An examination of the codon translation table (page 389) shows that approximately 25 percent of all random base-pair changes would be synonymous, giving an alternative codon for the same amino acid, while 75 percent of random changes will change the amino acid coded. For example, a change from AAT to AAC still codes for asparagine, but a change to ATT, ACT, AAA, AAG, AGT, TAT, CAT or GAT, all single-base-pair changes from AAT, changes the amino acid coded. So, if mutations of base pairs are at random and if the substitution of an amino acid made no difference to function, we would expect a 3:1 ratio of amino acid replacement to silent polymorphisms. The actual ratios found in *Drosophila* vary from 2:1 to 1:10. Clearly, there is a great excess of synonymous polymor-

phism, showing that most amino acid changes are subject to natural selection.

Message Within species there is great genetic variation. This is manifest at the morphological level of chromosome form and number and at the level of DNA segments that may have no observable developmental effects.

Variation within and among Populations

The various examples just given show that there are genetic differences among individuals within a population and also that the allelic frequencies differ among populations. The relative amounts of variation within and among populations vary from species to species, depending on history and environment. In humans, some gene frequencies (for example, those for skin color or hair form) obviously are well differentiated among populations and major geographical groups (so-called geographical races). If, however, we look at single structural genes identified immunologically or by electrophoresis rather than these outward phenotypic characters, the situation is rather different. Table 25-8 gives the allelic frequencies in a random sample of enzyme loci (chosen only for experimental convenience) in a European and an African population. Except for phosphoglucomutase-3, where the allelic frequencies are reversed between the populations, blacks and whites are very similar in their allelic distributions. In contrast to this random sample of loci, Table 25-9 shows the three loci for which Caucasians, Negroids, and Mongoloids are known to be most different from each other (Duffy and Rhesus blood groups and the P antigen) compared with the three polymorphic loci for which the races are most similar (Auberger blood group and Xg and secretor factors). Even for the most divergent loci, no race is homozygous for one allele that is absent in the other two races.

In general, different human populations show rather similar frequencies for polymorphic genes. Figure 25-5 is a **triallelic diagram** for the three main allelic classes I^A, I^B, and i of the ABO blood group. Each point represents the allelic composition of a population, where the three

Table 25-9. Examples of Extreme Differentiation and Close Similarity in Blood-Group Allelic Frequencies in Three Racial Groups

		Population		
Gene	Allele	Caucasoid	Negroid	Mongoloid
Duffy	*Fy*	0.0300	0.9393	0.0985
	Fy^d	0.4208	0.0607	0.9015
	Fy^b	0.5492	0.0000	0.0000
Rhesus	R_0	0.0186	0.7395	0.0409
	R_1	0.4036	0.0256	0.7591
	R_2	0.1670	0.0427	0.1951
	r	0.3820	0.1184	0.0049
	r′	0.0049	0.0707	0.0000
	Others	0.0239	0.0021	0.0000
P	P_1	0.5161	0.8911	0.1677
	P_2	0.4839	0.1089	0.8323
Auberger	Au^a	0.6213	0.6419	No data
	Au	0.3787	0.3581	No data
Xg	Xg^a	0.67	0.55	0.54
	Xg	0.33	0.45	0.46
Secretor	*Se*	0.5233	0.5727	No data
	se	0.4767	0.4273	No data

SOURCE: A summary is provided in L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations.* W. H. Freeman and Company, 1971, pp. 724–731. See this source for loci and for data sources.

Table 25-8. Allelic Frequencies at Seven Polymorphic Loci in Europeans and Black Africans

	Europeans			Africans		
Locus	Allele 1	Allele 2	Allele 3	Allele 1	Allele 2	Allele 3
Red-cell acid phosphatase	0.36	0.60	0.04	0.17	0.83	0.00
Phosphoglucomutase-1	0.77	0.23	0.00	0.79	0.21	0.00
Phosphoglucomutase-3	0.74	0.26	0.00	0.37	0.63	0.00
Adenylate kinase	0.95	0.05	0.00	1.00	0.00	0.00
Peptidase A	0.76	0.00	0.24	0.90	0.10	0.00
Peptidase D	0.99	0.01	0.00	0.95	0.03	0.02
Adenosine deaminase	0.94	0.06	0.00	0.97	0.03	0.00

SOURCE: R. C. Lewontin, *The Genetic Basis of Evolutionary Change.* Columbia University Press, 1974. Adapted from H. Harris, *The Principles of Human Biochemical Genetics.* North Holland, Amsterdam and London, 1970.

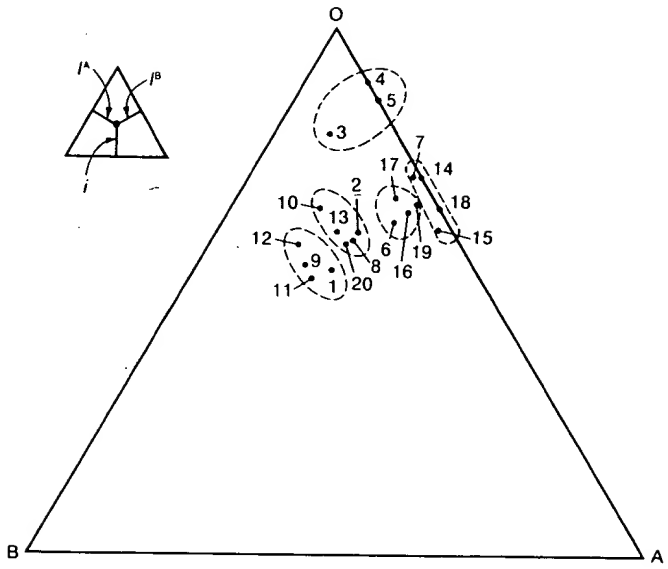


Figure 25-5 Triallelic diagram of the ABO blood-group allelic frequencies for human populations. Each point represents a population; the perpendicular distances from the point to the sides represent the allelic frequencies, as indicated in the small triangle. Populations 1 to 3 are African, 4 to 7 are American Indian, 8 to 13 are Asian, 14 to 15 are Australian aborigine, and 16 to 20 are European. Dashed lines enclose arbitrary classes with similar gene frequencies; these groupings do not correspond to "racial" classes. (After A. Jacquard, *Structures génétiques des populations.* Copyright 1970 by Masson et Cie.)

allelic frequencies can be read by taking the lengths of the perpendiculars from each side to the point. The diagram shows that all human populations are bunched together in the region of high i, intermediate I^A, and low I^B frequencies. Moreover, neighboring points (enclosed by dashed lines) do not correspond to geographical races, so that such races cannot be distinguished from each other by characteristic allelic frequencies for this gene. The study of polymorphic blood groups and enzyme loci in a variety of human populations has shown that about 85 percent of total human genetic diversity is found within local populations, about 8 percent is found among local populations within major geographical races, and the remaining 7 percent is found among the major geographical races. Clearly, the genes influencing skin color, hair form, and facial form that are well differentiated among races are not a random sample of structural gene loci.

Message In general, the genetic variation among individuals within human races is much greater than the average variation between races.

Quantitative Variation

The variation in quantitative characters cannot be described in terms of allelic frequencies, because individual loci and their alleles cannot be identified. Such variation can be characterized, however, by the amount of genetic variance (or the heritability of the trait) in the population. Figure 24-20 shows that many morphological and physiological traits in poultry have genetic variances of different amounts in different populations. A simple technique for estimating the additive genetic variance (page 730) of a character is to choose two groups of parents that are extremely different and then measure the difference between their offspring groups. The difference between offspring groups divided by the difference between parental groups is a measure of the heritability (h^2). Where this technique has been applied to morphological variation in *Drosophila,* for example, virtually every variable trait is found to have some genetic variance, so evolution of the trait can occur. Indeed, the method of estimating heritability just described is itself a kind of one-generation artificial-selection experiment.

It should not be supposed that all variable traits are heritable, however. Certain metabolic traits (such as resistance to high salt concentrations in *Drosophila*) show individual variation but no heritability. In general, behavioral traits have lower heritabilities than morphological traits, especially in organisms with more complex nervous systems that exhibit immense individual flexibility in central nervous states. Before any judgment can be made about the evolution of a particular quantitative trait, it is essential to determine if there is genetic variance for it in the population whose evolution is to be predicted. Thus, suggestions that such traits in the human species as performance on IQ tests, temperament, and social organization are in the process of evolving or have evolved at particular epochs in human history depend critically on evidence about whether there is genetic variation for these traits.

One of the most important findings in evolutionary genetics has been the discovery of substantial genetic variation underlying characters that show no morphological variation! These are called **canalized characters,** because the final outcome of their development is held within narrow bounds despite disturbing forces. Development is such that all the different genotypes for canalized characters have the same constant phenotype over the range of environments that is usual for the species. The genetic differences are revealed if the organisms are put in a stressful environment or if a severe mutation stresses the developmental system. For example, all wild-type *Drosophila* have exactly four scutellar bristles (Figure 25-6). If the recessive mutant *scute* is present, the number of bristles is reduced, but, in addition, there is variation from fly to fly. This variation is heritable, and lines with zero or one bristle and lines with three or four bristles can

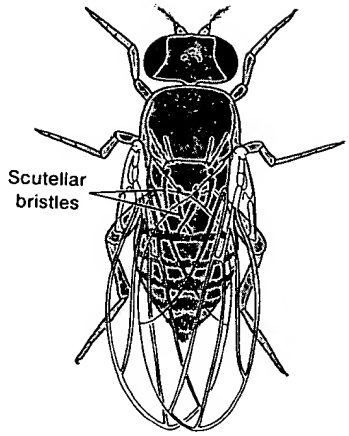


Figure 25-6 The scutellar bristles of the adult *Drosophila*, shown in blue. This is an example of a canalized character; all wild-type *Drosophila* will have four scutellar bristles in a very wide range of environments.

be obtained by selection in the presence of the *scute* mutation. When the mutation is removed, these lines now have two and six bristles, respectively. Similar experiments have been performed using extremely stressful environments in place of mutants. A consequence of such hidden genetic variation is that a character that is phenotypically uniform in a species may nevertheless undergo rapid evolution if a stressful environment uncovers the genetic variation.

Message Even characters with no apparent phenotypic variance may evolve when developmental conditions are changed drastically if genetic variation for the character is hidden by developmental canalization.

The Sources of Variation

The variational theory of evolution has a peculiar self-defeating property. If evolution occurs by the differential reproduction of different variants, we expect the variant with the highest rate of reproduction eventually to take over the population and all other genotypes to disappear. But then there is no longer any variation for further evolution. Genetic variation is the fuel for the evolutionary process, but differential reproduction consumes that fuel and so destroys the very condition necessary for further evolution. The possibility of continued evolution therefore is critically dependent on renewed variation.

For a given population, there are three sources of variation: mutation, recombination, and immigration of genes. However, recombination by itself does not produce variation unless alleles are segregating already at different loci; otherwise there is nothing to recombine. Similarly, immigration cannot provide variation if the entire species is homozygous for the same allele. Ultimately, the source of all variation must be mutation.

Variation from Mutations

Mutations are the *source* of variation, but the *process* of mutation does not itself drive evolution. The rate of change in gene frequency from the mutation process is very low because spontaneous mutation rates are low (Table 25-10). Let μ be the **mutation rate** from allele A to some other allele a (the probability that a gene copy **A** will become **a** during meiosis). If p_t is the frequency of the A allele in generation t, if $q_t = 1 - p_t$ is the frequency of the a allele, and if there are no other causes of gene-frequency change (no natural selection, for example), then the change in allelic frequency in one generation is

$$\Delta p = p_t - p_{t-1} = -\mu p_{t-1}$$

where p_{t-1} is the frequency in the preceding generation. This tells us that the frequency of A decreases (and the frequency of a increases) by an amount that is proportional to the mutation rate μ and to the proportion p of all the genes that are still available to mutate. Thus Δp gets smaller as the frequency of p itself decreases, because there are fewer and fewer A alleles to turn into a alleles. We can make the approximation that after n generations of mutation

$$p_n = p_0 e^{-n\mu}$$

where e is the base of the natural logarithms. This relation of gene frequency to number of generations is shown in Figure 25-7 for $\mu = 10^{-5}$. After 10,000 generations of continued mutation of A to a

$$p = p_0 e^{-(10^4)(10^{-5})} = p_0 e^{0.1} = 0.904\ p_0$$

So, if the population starts with only A alleles ($p_0 = 1.0$), it would still have only 10 percent a alleles after 10,000

Table 25-10. Some Point-Mutation Rates in Different Organisms

Organism	Gene	Mutation rate per generation
Bacteriophage	Host range	2.5×10^{-9}
Escherichia coli	Phage resistance	2×10^{-8}
Zea mays (corn)	*R* (color factor)	2.9×10^{-4}
	Y (yellow seeds)	2×10^{-6}
Drosophila melanogaster	Average lethal	2.6×10^{-5}

SOURCE: T. Dobzhansky, *Genetics and the Origin of Species*, 3d ed., revised. Columbia University Press, 1951.

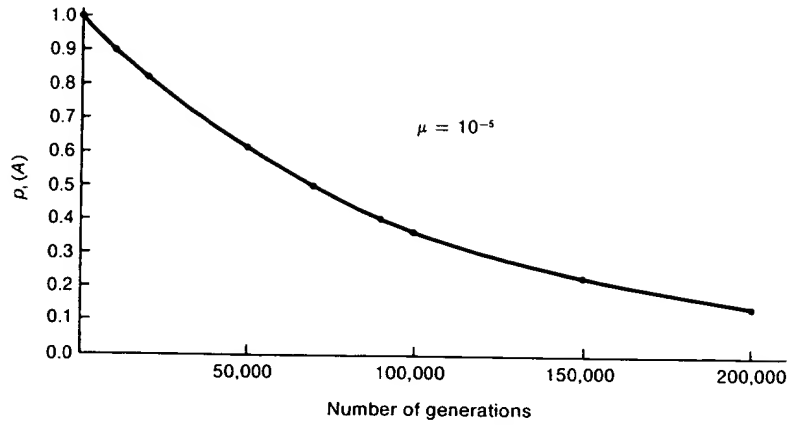


Figure 25-7 The change over generations in the frequency of a gene A due to mutation from A to a at a constant mutation rate (μ) of 10^{-5}.

generations at a rather high mutation rate and would require 60,000 additional generations to reduce p to 0.5. Even if mutation rates were doubled (say, by environmental mutagens), the rate of evolution would be very slow. For example, radiation levels of sufficient intensity to double the mutation rate over the reproductive lifetime of an individual human would be more than the amount necessary to cause death, so rapid genetic change in the species would not be one of the effects of increased radiation. Although we have many things to fear from environmental radiation pollution, turning into a species of monsters is not one of them.

If we look at the mutation process from the standpoint of the increase of a particular new allele rather than the decrease of the old form, the process is even slower. Most mutation rates that have been determined are the sum of all mutations of A to any mutant form with a detectable effect. Any *specific* base substitution is likely to be at least two orders of magnitude lower in frequency than the sum of all changes. So, precise reverse mutations ("back mutations") to the original allele A are unlikely, although many mutations may produce alleles that are *phenotypically* similar to the original.

Message Mutation rates are so low that mutation alone cannot account for the rapid evolution of populations and species.

It is not possible to measure locus-specific mutation rates for continuously varying characters, but the rate of accumulation of genetic variance can be determined. Beginning with a completely homozygous line of *Drosophila* derived from a natural population, $\frac{1}{1000}$ to $\frac{1}{500}$ of the genetic variance in bristle number in the original population is restored each generation by spontaneous mutation.

Variation from Recombination

The creation of genetic variance by recombination can be a much faster process than its creation by mutation. When two chromosomes of *Drosophila* with "normal" survival are allowed to recombine for a single generation, they produce an array of chromosomes that have 25 to 75 percent as much genetic variance in survival as the original wild population from which the parent chromosomes were sampled. This is simply a consequence of the fact that a single homologous pair of chromosomes that is heterozygous at n loci (taking into account only single and double crossovers) can produce $n(n-1)/2$ new unique gametic types from one generation of recombination. If the heterozygous loci are well spread out on the chromosomes, these new gametic types will be frequent and a considerable variance will be generated. Asexual organisms or organisms like bacteria that very seldom undergo sexual recombination do not have this source of variation, so that new mutations are the only way in which a change in gene combinations can be achieved. As a result, asexual organisms may evolve more slowly under natural selection than sexual organisms.

Variation from Migration

A further source of variation is migration into a population from other populations with different gene frequencies. If p_t is the frequency of an allele in the recipient population in generation t and P is the allelic frequency in a donor population (or the average over several donor populations), and if m is the frequency of migrants in the recipient population, then the gene frequency in the recipient population in the next generation is the result of mixing $(1-m)$ genes from the recipient with m genes from the donor population. Thus

$$p_{t+1} = (1-m)p_t + mP = p_t + m(P - p_t)$$

and

$$\Delta p = p_{t+1} - p_t = m(P - p_t)$$

The change in gene frequency is proportional to the difference in frequency between the recipient population

and the average of the donor populations. Unlike the mutation rate, the migration rate (m) can be large, so the change in frequency may be substantial.

We must understand *migration* as meaning any form of the introduction of genes from one population into another. So, for example, genes from Europeans have "migrated" into the population of African origin in North America steadily since the Africans were introduced as slaves. We can determine the amount of this migration by looking at the frequency of an allele that is found only in Europeans and not in Africans and comparing its frequency among blacks in North America.

We can use the formula for the change in gene frequency from migration if we modify it slightly to account for the fact that several generations of admixture have taken place. If the rate of admixture has not been too great, then (to a close order of approximation) the sum of the single-generation migration rates over several generations (let's call this M) will be related to the total change in the recipient population after these several generations by the same expression as the one used for changes due to migration. If, as before, P is the allelic frequency in the donor population and p_0 is the original frequency among the recipients, then

$$\Delta p_{\text{total}} = M(P - p_0)$$

so

$$M = \frac{\Delta p_{\text{total}}}{P - p_0}$$

For example, the Duffy blood-group allele Fy^a is absent in Africa but has a frequency of 0.42 in whites from the state of Georgia. Among blacks from Georgia, the Fy^a frequency is 0.046. Therefore, the total migration of genes from whites into the black population since the introduction of slaves in the eighteenth century is

$$M = \frac{\Delta p_{\text{total}}}{P - p} = \frac{0.046 - 0}{0.42 - 0} = 0.1095$$

When the same analysis is carried out on American blacks from Oakland (California) and Detroit, M is 0.22 and 0.26, respectively, showing either greater admixture rates in these cities than in Georgia or differential movement into these cities by American blacks who have more European ancestry. In any case, the genetic variation at the Fy locus has been increased by this admixture.

The Origin of New Functions

Point mutations or chromosomal rearrangements are themselves a limited source of variation for evolution because they can only alter a function or change one kind of function into another. To add quite new functions requires expansion in the total repertoire of genes through duplication and polyploidy, followed by a divergence between the duplicated genes, presumably by the usual process of mutation. Expansion of the genome by polyploidy has clearly been a frequent process, at least in plants. Figure 25-8 shows the frequency distribution of haploid chromosome numbers among dicotyledonous plant species. Even numbers are much more common than odd numbers—a consequence of frequent polyploidy.

Once an expansion of the total DNA of the genome has occurred, it may require only a few base substitutions in a gene to provide it with a new function. For example, B. Hall has experimentally changed a gene to a new function in *Escherichia coli.* In addition to the *lac Z* genes specifying the usual lactose-fermenting β-galactosidase activity in *E. coli,* another structural gene locus *ebg* specifies another β-galactosidase that does not ferment lactose, although it is induced by lactose. The natural function of this second enzyme is unknown. Hall was able to alter this gene into one specifying an enzyme that ferments another substrate, lactobionate. To do so, it was necessary to alter the regulatory element to a constitutive state and to produce three successive structural-gene mutations.

Message Evolution would come to a stop by running out of variation if new genetic variation were not added to populations by mutation, recombination, and migration. Ultimately, all new variation is derived from gene and chromosome mutations.

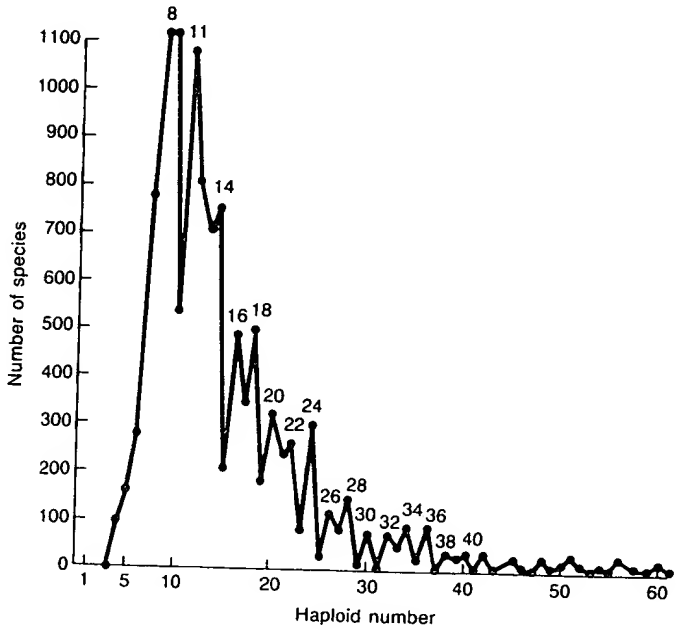


Figure 25-8 Frequency distribution of haploid chromosome numbers in dicotyledonous plants. (From Verne Grant, *The Origin of Adaptation.* Copyright 1963 by Columbia University Press.)

The Effect of Sexual Reproduction on Variation

The evolutionary theorists of the nineteenth century encountered a fundamental difficulty in dealing with Darwin's theory of evolution through natural selection. The possibility of continued evolution by natural selection is limited by the amount of genetic variation. But biologists of the nineteenth century, including Darwin, believed in one form or another of **blending inheritance,** a model postulating that the characteristics of each offspring are some intermediate mixture of the parental characters. Such a model of inheritance has fatal implications for a theory of evolution that depends on variation.

Suppose that some trait (say, height) has a distribution in the population and that individuals mate more or less at random. If intermediate individuals mated with each other, they would produce only intermediate offspring according to a blending model. The mating of a tall with a short individual also would produce only intermediate offspring. Only the mating of tall with tall individuals and short with short individuals would preserve extreme types. The net result of all matings would be an increase in intermediate types and a decrease in extreme types. The variance of the distribution would shrink, simply as a result of sexual reproduction. In fact, it can be shown that the variance is *cut in half* in each generation, so that the population would be essentially uniformly intermediate in height before very many generations had passed. There then would be no variation on which natural selection could operate. This was a very serious problem for the early Darwinists; it made it necessary for Darwin to assume that new variation is generated at a very rapid rate by the inheritance of characters acquired by individuals during their lifetimes.

The rediscovery of Mendelism changed this picture completely. Because of the discrete nature of the Mendelian genes and the segregation of alleles at meiosis, a cross of intermediate with intermediate individuals does *not* result in all intermediate offspring. On the contrary, extreme types (homozygotes) segregate out of the cross. To see the consequence of Mendelian inheritance for genetic variation, consider a population in which males and females mate with each other at random with respect to some gene locus A; that is, individuals do not choose their mates preferentially with respect to the partial genotype at the locus. Such random mating is equivalent to mixing all the sperm and all the eggs in the population together and then matching randomly drawn sperm with randomly drawn eggs.

If the frequency of allele A is p in both the sperm and the eggs and the frequency of allele a is $q = 1 - p$, then the consequences of random unions of sperm and eggs are shown in Figure 25-9. The probability that both the sperm and the egg will carry A is $p \times p = p^2$, so this will

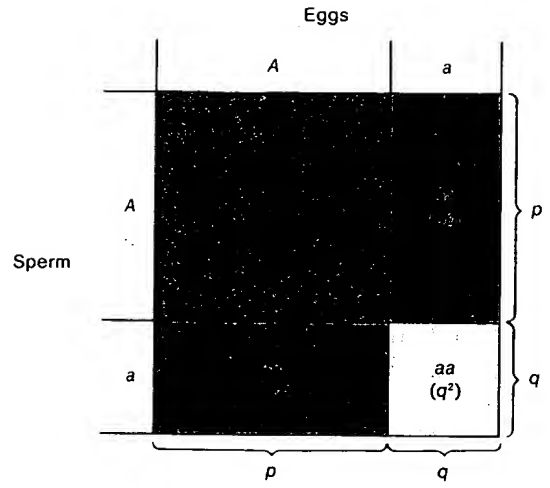


Figure 25-9 The Hardy-Weinberg equilibrium frequencies that result from random mating. The frequencies of A and a among both eggs and sperm are p and q $(= 1 - p)$, respectively. The total frequencies of the zygote genotypes are p^2 for AA, $2pq$ for Aa, and q^2 for aa. The frequency of the allele A in the zygotes is the frequency of AA plus half the frequency of Aa, or $p^2 + pq = p(p + q) = p$.

be the frequency of AA homozygotes in the next generation. In like manner, the chance of heterozygotes Aa will be $(p \times q) + (q \times p) = 2pq$, and the chance of homozygotes aa will be $q \times q = q^2$. The three genotypes, after a generation of random mating, will be in the frequencies $p^2{:}2pq{:}q^2$. As the figure shows, the allelic frequency of A has not changed and is still p. Therefore, in the second generation, the frequencies of the three genotypes will again be $p^2{:}2pq{:}q^2$, and so on, forever.

Message Mendelian segregation has the property that random mating results in an equilibrium distribution of genotypes after only one generation, so that genetic variation is maintained.

The equilibrium distribution

AA	Aa	aa
p^2	$2pq$	q^2

is called the **Hardy-Weinberg equilibrium** after those who independently discovered it. (A third independent discovery was made by the Russian geneticist Sergei Tschetverikov.)

The Hardy-Weinberg equilibrium means that sexual reproduction does not cause a constant reduction in genetic variation in each generation; on the contrary, the amount of variation remains constant generation after generation, in the absence of other disturbing forces. The equilibrium is the direct consequence of the segregation of alleles at meiosis in heterozygotes.

Numerically, the equilibrium shows that irrespective of the particular mixture of genotypes in the parental generation, the genotypic distribution after one round of mating is completely specified by the allelic frequency p. For example, consider three hypothetical populations, all having the same frequency of $A(p = 0.3)$:

	AA	Aa	aa
I	0.3	0.0	0.7
II	0.2	0.2	0.6
III	0.1	0.4	0.5

After one generation of random mating, each of the three populations will have the same genotypic frequencies:

AA	Aa	aa
$(0.3)^2 = 0.09$	$2(0.3)(0.7) = 0.42$	$(0.7)^2 = 0.49$

and they will remain so indefinitely.

One consequence of the Hardy-Weinberg proportions is that rare alleles are virtually never in homozygous condition. An allele with a frequency of 0.001 occurs in homozygotes at a frequency of only one in a million; most copies of such rare alleles are found in heterozygotes. In general, since two copies of an allele are in homozygotes but only one copy of that allele is in each heterozygote, the relative frequency of the allele in heterozygotes (as opposed to homozygotes) is

$$\frac{2pq}{2q^2} = \frac{p}{q}$$

which for $q = 0.001$ is a ratio of 999:1. The general relation between homozygote and heterozygote frequencies as a function of allele frequencies is shown in Figure 25-10.

In our derivation of the equilibrium, we assumed that the allelic frequency p is the same in sperm and eggs. The Hardy-Weinberg equilibrium theorem does not apply to sex-linked genes if males and females start with unequal gene frequencies (see Problem 7 at the end of the chapter).

The Hardy-Weinberg equilibrium was derived on the assumption of "random mating," but we must carefully distinguish two meanings of that process. First, we may mean that individuals do not choose their mates on the basis of some heritable character. Human beings are random-mating with respect to blood groups in this first sense, because they generally do not know the blood type of their prospective mates, and even if they did, it is unlikely that blood type would be used as a criterion for choice. In the first sense, random mating will occur with respect to genes that have no effect on appearance, behavior, smell, or other characteristics that directly influence mate choice.

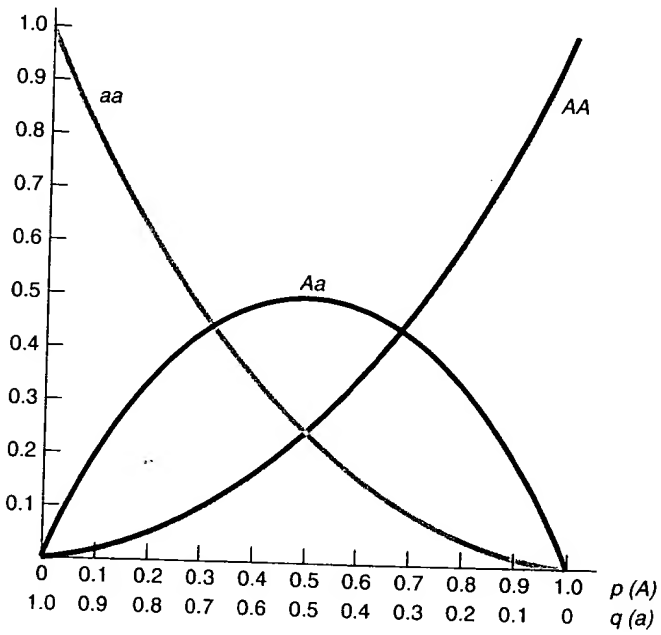


Figure 25-10 Curves showing the proportions of homozygotes AA (blue line), homozygotes aa (yellow line), and heterozygotes Aa (green line) in populations of different allelic frequencies if the populations are in Hardy-Weinberg equilibrium.

There is a second sense of random mating that is relevant when there is any subdivision of a species into subgroups. If there is genetic differentiation between subgroups so that the frequencies of alleles differ from group to group and if individuals tend to mate within their own subgroup **(endogamy),** then with respect to the species as a whole, mating is not at random and frequencies of genotypes will depart more or less from Hardy-Weinberg frequencies. In this sense, human beings are not random-mating, because ethnic and racial groups differ from each other in gene frequencies and people show high rates of endogamy, not only within major races but also within local ethnic groups. Spaniards and Russians differ in their ABO blood group frequencies, Spaniards marry Spaniards and Russians marry Russians, so there is unintentional nonrandom mating with respect to ABO blood groups. Table 25-11 shows random mating in the first sense and nonrandom mating in the second sense for the MN blood group. Within Eskimos, Egyptian, Chinese, and Australian subpopulations, females do not choose their mates by MN type, and, thus, Hardy-Weinberg equilibrium exists *within* the subpopulations. But Egyptians do not mate with Eskimos or Australian aborigines, so the nonrandom associations in the human species *as a whole* result in large differences in genotype frequencies and departure from Hardy-Weinberg equilibrium.

Table 25-11. Comparison between Observed Frequencies of Genotypes for the MN Blood-Group Locus and the Frequencies Expected from Random Mating

	Observed			Expected		
Population	*MM*	*MN*	*NN*	*MM*	*MN*	*NN*
Eskimo	0.835	0.156	0.009	0.834	0.159	0.008
Egyptian	0.278	0.489	0.233	0.274	0.499	0.228
Chinese	0.332	0.486	0.182	0.331	0.488	0.181
Australian aborigine	0.024	0.304	0.672	0.031	0.290	0.679

NOTE: The expected frequencies are computed according to the Hardy-Weinberg equilibrium, using the values of p and q computed from the observed frequencies.

Inbreeding and Assortative Mating

Random mating with respect to a locus is common, but it is not universal. Two kinds of deviation from random mating must be distinguished. First, individuals may mate with each other nonrandomly either because of their degree of common ancestry or their degree of genetic relationship. If mating between relatives occurs more commonly than would occur by pure chance, then the population is **inbreeding.** If mating between relatives is less common than would occur by chance, then the population is said to be undergoing **enforced outbreeding,** or **negative inbreeding.**

Second, individuals may tend to choose each other as mates, not because of their degree of genetic relationship but because of their degree of resemblance to each other at some locus. Bias toward mating of like with like is called **positive assortative mating.** Mating with unlike partners is called **negative assortative mating.** Assortative mating is never complete.

Inbreeding levels in natural populations are a consequence of geographical distribution, of the mechanism of reproduction, and of behavioral characteristics. If close relatives occupy adjacent areas, then simple proximity may result in inbreeding. The seeds of many plants, for example, fall very close to the parental source and the pollen is not widely spread, so a high frequency of sib mating occurs. Some plants (such as corn) can be self-pollinated as well as cross-pollinated, so that wind pollination results in some very close inbreeding. Yet other plants, like the peanut, are obligatorily selfed. Many small mammals (such as house mice) live and mate in restricted family groups that persist generation after generation. Humans, on the other hand, generally have complex mating taboos and proscriptions that reduce inbreeding.

Assortative mating for some traits is common. In humans, there is a positive assortative mating bias for skin color and height, for example. An important difference between assortative mating and inbreeding is that the former is specific to a trait whereas the latter applies to the entire genome. Individuals may mate assortatively with respect to height but at random with respect to blood group. Cousins, on the other hand, resemble each other genetically on the average to the same degree at all loci.

For both positive assortative mating and inbreeding, the consequence to population structure is the same: there is an increase in homozygosity above the level predicted by the Hardy-Weinberg equilibrium. If two individuals are related, they have at least one common ancestor. Thus, there is some chance that an allele carried by one of them and an allele carried by the other are both descended from the identical DNA molecule. The result is that there is an extra chance of **homozygosity by descent,** to be added to the chance of homozygosity ($p^2 + q^2$) that arises from the random mating of unrelated individuals. The probability of homozygosity by descent is called the **inbreeding coefficient (F).** Figure 25-11 illustrates the calculation of the probability of homozygosity by descent. Individuals I and II are full sibs because they share both parents. We label each allele in the parents uniquely to keep track of them. Individuals I and II mate to produce individual III. If individual I is A_1A_3 and the gamete that it contributes to III contains the allele A_1, then we would like to calculate the probability that the gamete produced by II is also A_1. The chance is $\frac{1}{2}$ that II will receive A_1 from its father, and if it does, the chance is $\frac{1}{2}$ that II will pass A_1 on to the gamete in question. Thus, the probability that III will receive an A_1 from II is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, and this is the chance that III — the product of a full-sib mating — will be homozygous by descent.

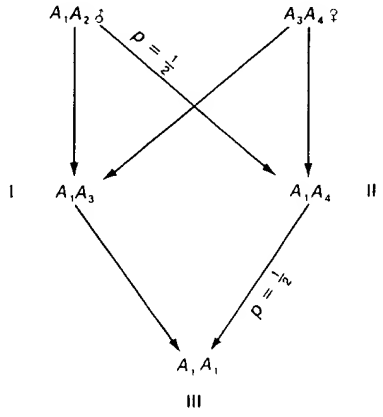


Figure 25-11 Calculation of homozygosity by descent for an offspring (III) of a brother-sister (I-II) mating. The probability that II will receive A_1 from its father is $\frac{1}{2}$; if it does, the probability that II will pass A_1 on to the generation producing III is $\frac{1}{2}$. Thus, the probability that III will receive an A_1 from II is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Related Proceedings Appendix

None